Protecting Students' Privacy in the Age of MOOCs

Drew Bent, Nitah Onsongo, Andy Trattner

Massachusetts Institute of Technology

Foundations of Information Policy (6.805)

December 8, 2017

{ bent, nonsongo, trattner } @ mit.edu

Executive Summary

The growth of the internet has enabled millions to access high quality education through Massive Open Online Courses (MOOCs). In turn, enormous student populations and modern data science techniques provide researchers with unprecedented opportunities to study the human learning experience. While these factors unquestionably bequeath great social benefits, they also heighten student privacy concerns beyond the scope of existing legislation in the United States. In order to protect the privacy of students, we propose a Student Internet Privacy Act that regulates the activities of MOOC providers without undermining their social benefits.

The approach of this paper is twofold. First, we return to the basics in order to understand MOOCs, their benefits, and their growing privacy concerns. Second, we address the legislative landscape, considering the current responses to such privacy concerns and their limitations, before ultimately proposing our own framework for a new federal law.

We find that the benefits of MOOCs—which include edX, Coursera, and Udacity—largely fall into the categories of access and academic research. With MOOCs, users have access not only to new, high-quality educational material, but also to new degrees and job opportunities at lower costs than previously available. In terms of academic research, the big data that MOOCs collect allow for publication of novel insights ranging from macro-level international comparisons to micro-level behavioral analyses.

With such benefits also come privacy concerns, and we study them under several contexts. The business models of MOOC providers are varied, and ones such as job matching and targeted advertising have the potential to disclose user data in unexpected ways that go well beyond educational purposes. Additionally, research taking place inside a company opens up the risk of personnel having access to sensitive user data without being governed by institutional review boards. Finally, across their entire platforms, MOOC providers severely lack the data transparency and control to which users have a right. The case of the data-management company, inBloom, exemplifies the dangers of not providing users with stronger data control. Parents and schools were so upset with inBloom's vast and uncontrolled collection of sensitive data that the company was forced to shut down in 2014.

Given the inadequacies of current federal legislation like the Family Educational Rights and Privacy Act, states have enacted dozens of their own laws to prohibit commercial uses of student data. California has led the way with its Student Online Personal Information Protection Act (SOPIPA). However, these laws fall short in several important regards: they don't cover students nationally, they concern only K–12 students and exclude MOOC learners, they fail to guarantee students with transparency and control over their data, and they often lack strong provisions for enforcement and redress.

In order to close these gaps in existing legislation, we propose a new federal law to protect the privacy of learners on both K–12 platforms and MOOCs nationwide. We first state four key principles that underlie student privacy—Transparency, Control of Data, Focused Collection, and Educational Use and Context—drawing upon a number of well grounded privacy notions that stem from the Electronic Privacy Information Center's privacy principles, President Barack Obama's Consumer Privacy Bill of Rights, and the Organisation for Economic Co-operation and Development's Privacy Framework. Together, these principles embody the rights of students in a broad range of settings and not just schools.

We then outline the key definitions and provisions that could enshrine these principles in law. While some definitions such as "personally identifiable information" are adopted from California's SOPIPA, others such as "operators" are rewritten to broaden the coverage of our bill. We notably extend our protections to all users of MOOCs and K–12 websites—not just those enrolled in traditional schools—and we classify all these learners as "students."

The proposed bill would prohibit MOOC and K–12 educational technology providers from engaging in exploitative practices, including targeted advertising and the sale of student data. The bill would also require MOOC providers to offer data dashboards to students for increased transparency, as well as the option to delete, amend, and export one's own data. The bill would minimize chilling effects on MOOC business models, and it would not prevent academic research from continuing. For researchers operating within a MOOC provider, new access control requirements would be put in place to ensure proper care is taken with sensitive student data. The proposed bill also includes strong enforcement provisions through both civil enforcement action by the Federal Trade Commission and United States Attorney General, as well as a private right of action.

If enacted, the Student Internet Privacy Act could usher in a new era for student privacy. We end by considering the implications such a bill would have on various stakeholders and its prospects for becoming law.

Contents

1. Background

Online privacy is an increasingly relevant issue as new Massive Open Online Courses (MOOCs) and educational technology (edtech) websites emerge. The combination of large amounts of user data and increased processing power allows MOOC providers to draw more unwanted inferences about users than previously possible. Data can also be shared more readily without users' permission.

In the decade since MOOCs began in 2008, the technological landscape has noticeably shifted. Whereas the first MOOCs were experiments that a brave few opted into, the MOOCs of today form a key part of millions of students' educations. MOOC providers are experimenting with new business models and are even offering official university credit for some courses.[1]

Yet despite the integration of MOOCs into mainstream education, student privacy remains an afterthought due to a slow-moving policy landscape. The prevailing federal law, the Family Educational Rights and Privacy Act, has only been updated twice in the past decade, and in neither instance was the advent of MOOCs or new privacy concerns addressed.[2] As a result, MOOC providers remain unchecked in their data collection, processing, and dissemination practices.

In order to address the privacy risks posed by MOOCs, we first need a firm grasp of privacy and its nuances. In this section, we begin by considering the foundations of privacy in the United States. We then consider the history of educational privacy. We finish the section with a discussion of the MOOC phenomenon and how it pertains to privacy, to which we tune the scope of the paper and our recommendations.

1.1 What is Privacy?

There is no one precise definition of privacy, which explains why privacy law in the United States has largely evolved through a haphazard and case-by-case basis. Even the definitions that do exist are broad and not readily applicable. The Oxford English Dictionary, for example, defines privacy in general terms as the "state or condition of being free from being observed or disturbed by other people."[3]

Academics have tried to understand privacy by explaining the sociological motivations for protecting it. For instance, in Daniel Solove's seminal "Taxonomy of Privacy,"[4] he writes, "Privacy is the relief from a range of kinds of social friction. It enables people to engage in

[1] Straumsheim, Carl, "MIT Deems MicroMasters a Success."
[2] Electronic Privacy Information Center, "Family Educational Rights and Privacy Act (FERPA)."
[3] "privacy, n.1". Oxford Living Dictionaries Online.
[4] Solove, Daniel J, "A Taxonomy of Privacy."

worthwhile activities in ways that they would otherwise find difficult or impossible."[5] In particular, he notes that a lack of privacy protection can result in "reputational injury,"[6] the behavioral chilling effects of surveillance,[7] and a systematic increase in the risk of future harm.[8] These sociological motivations suggest certain implications for privacy law. One example is Solove's observation that the very act of information collection—even without processing or dissemination—can lead to the aforementioned privacy problems, and thus should be carefully monitored.[9] However, such analyses of privacy rights still don't answer which particular types of information should be protected and to what extent.

Given the evolving nature of sociology and the challenges of establishing unambiguous privacy rights, it is unsurprising that protection of privacy in the United States has been largely reactionary. The U.S. Constitution did not originally explicate the right to privacy, so privacy rights have instead developed through amendments, subsequent interpretation, and case law.[10] As Solove and Woodrow Hartzog put it, "Privacy law in the United States has developed in a fragmented fashion and is currently a hodgepodge of various constitutional protections, federal and state statutes, torts, regulatory rules, and treaties."[11]

The result is that "privacy law in the United States is sectoral, with different laws regulating different industries and economic sectors."[12]

1.2 Educational Privacy in the United States

Given the sectoral nature of privacy law in the United States, we are interested in focusing our attention on how educational privacy developed in particular.

President Gerald Ford signed the Family Educational Rights and Privacy Act of 1974 (FERPA) into law on August 21, marking a watershed moment in U.S. educational privacy. Law Professor Mary Margaret Penrose points out that Ford acted within two weeks after Nixon's resignation,[13] writing that the Watergate scandal in 1972 was partially responsible for "[creating] a climate that gave rise to FERPA."[14] In particular, the "law's legislative history suggests the distrustful climate surrounding Watergate cascaded into other privacy areas, including education."[15]

Senator James Buckley originally introduced the bill because he saw "systematic violations of the privacy of students and parents by the schools through the unauthorized,

[5] Solove, Daniel J, "A Taxonomy of Privacy," 484.
[6] Ibid., 486.
[7] Ibid., 487.
[8] Ibid.
[9] Ibid., 491-499.
[10] e.g. freedom of association and privacy of membership in the First Amendment, NAACP v. Patterson 1958 (357 U.S. 449).
[11] Solove, Daniel J. and Woodrow Hartzog, "The FTC and the New Common Law of Privacy," 587.
[12] Ibid.
[13] Penrose, Mary Margaret, "In The Name Of Watergate: Returning FERPA To Its Original Design," 76.
[14] Ibid., 78.
[15] Ibid., 77.

inappropriate release of personal data to various individuals and organizations,"[16] which was concerning to voters at such a time when "parents...were granted less access to their children's complete school records than authorities like the law enforcement agencies and health departments."[17]

The resulting FERPA laid the groundwork for key principles in educational privacy, including the right to access and control one's own data. This manifested itself in requirements for educational institutions that receive funds under the Secretary or Department of Education. The statute required that schools grant students and parents the following rights: (1) right to control who can access their educational records; (2) right to assess and review their educational records; (3) right to request their educational records be updated in case of inaccuracies; and (4) right to report FERPA violations to the US Department of Education.[18]

FERPA established educational privacy rights in the United States. However, it was passed over 40 years ago yet remains the only significant body of federal legislation tailored towards privacy in education. We delve further into FERPA and its insufficiency in the digital age in Section 4, where we consider its enforcement (or lack thereof) and relevant amendments.

1.3 MOOCs and Privacy

In theoretical computer science, the famous halting problem has been shown to be undecidable—there is no algorithm that can predict if any arbitrary computer program will terminate on all given inputs. Likewise, seemingly deterministically, FERPA and privacy law in general have had trouble anticipating, and therefore coping with, the progress of technology. Existing frameworks simply cannot handle the new situation we find ourselves in today; their very structures are their downfall.

Since the enactment of FERPA, computing power and the internet have exponentially increased the capacity of organizations to collect, process, and disseminate information. Nowhere is this more clear than with MOOCs, which epitomize the duality of the technology-fueled information explosion. They have benefitted users enormously, but there has been a corresponding increase in privacy risks as well. As detailed in Section 3, certain business models and data-handling practices (or lack thereof) infringe upon user privacy when translated into the realm of education. Unfortunately, FERPA is not well-equipped to handle these issues as it covers schools but not companies. Legislative bodies are beginning to react; for instance, California prohibited targeted advertising and other practices in 2014. However, we will see in Section 4 that these solutions have their own problems.

[16] Jones, Meg Leta and Lucas Regner, "Users or Students? Privacy in University MOOCs," 1479.
[17] Ibid.
[18] Ibid., 1481.

Looking back 40 years at privacy awareness in the U.S., it seems that today's generation has had its Watergate moment with regards to online education, yet the analogous FERPA is still missing. The digitalization of the traditional classroom continues, but at present, fundamental privacy issues remain unresolved.

1.4 Scope of this Paper

MOOC and Edtech Providers

In this paper, we restrict our discussion of online learning services primarily to MOOC providers, because they are the services that have grown the quickest over the past decade while avoiding any regulation. In defining MOOC providers, we consider online services that serve the same primary functions as traditional schools, including courseware and some form of certification. This definition encompasses the Big Three in MOOCs—Coursera, edX, and Udacity—as well as others like Udemy and smaller regional providers. All of them include "courses," which Merriam-Webster defines as "a number of lectures or other matter dealing with a subject." They also include certifications for some of their courses, whether a proprietary credential or a degree affiliated with a partner institution. We also note that most of these MOOCs target students of higher education, as well as lifelong learners. However, in line with the prevailing view among experts,[19] this definition of MOOCs excludes many other edtech websites like StackExchange, Piazza, and Khan Academy. In the case of Khan Academy, at present, the content is organized into short videos and modules, not courses.

In our proposed legislation, however, we do still address this broader set of edtech providers, particularly K–12 ones. These include any websites that serve K–12 students for educational purposes, whether inside or outside a school setting. Khan Academy and Knewton are two examples. While our paper's unique contribution to the literature is the protection of user privacy on MOOCs, we build upon other legislation that already protects users on K–12 edtech websites. As such, our framework aims to protect users of both types: those on all K–12 edtech websites and those on MOOCs.

Students

We consider all users of such aforementioned MOOCs and K–12 edtech websites to be students. Therefore, they should all be afforded student privacy rights online. This group includes learners who are not students in the traditional sense and may not be enrolled in schools, such as retired engineers brushing up on differential equations with edX or managers learning to code on Coursera. Additionally, while we focus this paper on American learners, other international users of MOOCs would also benefit from our policy recommendations.

[19] Jon Daries, video conference with authors, 13 November 2017.

Our decision to develop a more encompassing definition of students is motivated by the principles underlying basic privacy rights. These rights are not exclusive to students enrolled in federally funded schools, although legislative constraints led to FERPA being designed that way. It is important to recognize that Solove's reputational injury and surveillance-induced behavioral effects can harm anyone who is in an educational setting.[20] Any learner who uses a service that serves the purpose of a school should be afforded the same privacy rights.

In summary, we believe that the internet era has expanded the circumstances under which anyone can pursue education, especially through MOOCs. Therefore, there should be a corresponding expansion in whom should be granted privacy rights.

1.5 Outline of Policy Recommendation

To combat privacy concerns posed by MOOCs' unregulated use of data, in this paper we propose a bill titled the Student Internet Privacy Act. The bill would prohibit MOOC and K–12 edtech providers from engaging in a variety of exploitative practices, as well as require them to grant specific privacy rights to students. The definitions and clauses, outlined below, will be developed in detail in Section 5.

Key Definitions

1. Operator
2. Student
3. Personally Identifiable Information
4. Educational Purposes
5. Targeted Advertising

Legislative Clauses

1. No targeted advertising
2. No student profiles except for educational uses
3. No sale of student information
4. Research exemption
5. Data dashboards
6. Deletion of one's own account
7. Amendment of one's own data
8. Exporting one's own data
9. Access control
10. Enforcement and redress

[20] Solove, Daniel J, "A Taxonomy of Privacy," 486-7.

2. Benefits of MOOCs

MOOCs offer unprecedented access to quality education as well as treasure troves of data for academic research. Both of these aspects include their own, unique benefits to society. Whereas ten years ago these benefits were speculative, the ensuing rise of MOOCs has shed light on their impressive and ongoing impact.

2.1 Access

MOOCs provide students with access to new opportunities and information in several ways. They are designed for scale and available across borders and languages. Top universities and professors around the globe produce content for dissemination, and millions of students of all ages and backgrounds have registered with MOOC providers.[21] Although gender, geographic, and socioeconomic imbalances still exist in learner populations, MOOCs are able to lower the higher education barrier to entry overall.

Similar to YouTube and other services that facilitate the sharing of content with massive quantities of users, MOOCs usually give away their content for students to consume freely. In so doing, MOOCs expand access to students who might otherwise fall outside of the traditional target audience range for college classes,[22] and they also provide opportunities for teachers to re-engage with material and pass it along to other learners.[23]

The secret to providing free content is to have a sustainable business model, which can benefit students. Users of YouTube are familiar with how the platform includes advertising and automatic, personalized suggestions. These features provide revenue and enable the company to offer its free services. It is natural to wonder if MOOCs will sustain their growth through similar business models. For example, Coursera is the largest MOOC provider,[24] having served over 25 million students. Coursera has achieved this scale through venture funding[25] as a for-profit corporation, but it is currently searching for a sustainable business model.

In recent years, MOOCs have increasingly experimented with one business model in particular: certification. While this means that MOOCs are not entirely free anymore, their ability to offer certification at a much lower cost than traditional universities means that they are increasing students' access to credentials in addition to content. This further helps

[21] Shah, Dhawal. 2017. "Massive List of MOOC Provides Around The World."
[22] Bayeck, R. Y., "Exploratory study of MOOC learners' demographics and motivation: The case of students involved in groups," 224.
[23] Seaton, Daniel et al., "Enrollment in MITx MOOCs: Are We Educating Educators?"
[24] Shah, Dhawal. 2017. "Massive List of MOOC Provides Around The World."
[25] Crunchbase, "Coursera."

professionals across the country and globe who may not otherwise be able to afford an expensive degree.

For instance, the MIT MicroMasters is a program that launched on edX in October 2016 and offers a new type of "micro" degree for students who take a series of online classes and pass a final exam.[26] The degree is not the same as a university graduate degree, but it comes at a fraction of the cost: around $600-1400.[27] It also carries weight: the university granting the MicroMaster will usually accept it as transfer credit if the student proceeds to the residential program.[28] Georgia Tech similarly began offering an online master's in computer science in 2014. The OMSCS program, as it's called, is an accredited degree yet a fraction of the cost of the residential degree.[29] Arizona State University partnered with edX in 2015 to do the same, and further expanded the scope to include the entire freshman year curriculum.[30] In all these cases, students are given newfound access to affordable credentials.

In addition to credentials, MOOCs are trying to increase students' access to jobs. Udacity is a pioneer in this job matching business model; the company believes "that the ultimate value proposition of education is employment" and the life to which it leads. Udacity has partnered with companies like Google, AT&T, and Accenture as part of its matching service.[31]

Between educational content, credentials, and jobs, MOOCs have increased access to a variety of opportunities that were previously unavailable to many students.

2.2 Research

In addition to access, the digital nature of MOOCs supports powerful educational research by allowing providers to collect information from student interactions in the platform. The extensive logging, coupled with massive participation, results in datasets of considerable size. This data can then be processed and analyzed both by academics at partner institutions and by employees within the MOOC providers. Already, MOOC research has yielded valuable insights on multiple levels, from macro international comparisons to micro behavioral analysis.

Research findings have shown new big-picture trends,[32] confirmed various pedagogical theories,[33] refuted the efficacy of certain traditional practices, and raised new questions

[26] Straumsheim, Carl, "MIT Deems MicroMasters a Success."
[27] Fischer, Jill, "Top 4 FAQ about the MicroMasters Credential."
[28] Ibid.
[29] Georgia Tech, "Why OMS CS?"
[30] Straumsheim, Carl, "MOOCs for (a Year's) Credit."
[31] Diaz-Hernandez, Ana, "Announcing Talent Source: Connecting Students & Employers With Our New Candidate Sourcing Tool."
[32] Chuang, Isaac and Andrew Ho, "HarvardX and MITx: Four Years of Open Online Courses -- Fall 2012-Summer 2016."
[33] Chen, Z. et al., "Researching for better instructional methods using AB experiments in MOOCs: results and challenges," 16.

about how students learn. For instance, Harvard and MIT published a four-year review that showed a strong correlation between the World Bank Human Development Index and certificates completed in a given country.[34] MIT's Office of Open Learning has published studies on how retrieval practice improves knowledge recall, how interleaving problem types results in "greater learning gains," and how worked examples are more helpful than solving new problems for non-experts. David Pritchard, a prominent MOOC researcher at MIT, coauthored a study on how "drag-and-drop problems seem to develop individual expert skills significantly better than multiple choice format."[35] While not all research requests achieve dramatic results like these,[36] the breakthrough insights discovered thus far promise a future ripe with opportunity.

Such research often originates from students' personal information. While there are some privacy concerns associated with this research, as will be discussed in the next section, any restrictions on using student data in academic settings could limit research findings and their resulting benefits. The de-identification of student data in particular would dampen these benefits of educational research.

There are two main reasons for de-identification affecting research results. The first issue is one of scope. Daries et al. warn that "at this point it may be possible to take for granted that any standard for de-identification [of student data] will increase over time."[37] Essentially, every year we are confronted with more data that can be cross-referenced with social media and other publicly available information to identify students. Therefore, as times goes, more data would become unavailable for research use if researchers could only use de-identified datasets. The scope of educational research would shrink.

The second issue is the tradeoff between available data and the quality of the resultant findings. The chilling effects of de-identification on education research include biased final findings, datasets that are not a true representation of the original data, and limited future analyses.[38] Daries et al. provide evidence that restricted access to personal information "would, at best, slow down the advancement of knowledge. At worst, these limits would prevent groundbreaking research from ever being conducted."[39]

In summary, a clear benefit of MOOCs is that they enhance educational research through rich, new datasets. However, it is important to note that these robust findings require complete datasets and as little de-identification as possible.

[34] Chuang, Isaac and Andrew Ho, "HarvardX and MITx: Four Years of Open Online Courses -- Fall 2012-Summer 2016."
[35] Chen, Z. et al., "Researching for better instructional methods using AB experiments in MOOCs: results and challenges," 16.
[36] Jon Daries, video conference with authors, 13 November 2017.
[37] Daries, Jon P. et al., "Privacy, Anonymity, and Big Data in the Social Sciences," 62.
[38] Ibid., 57.
[39] Ibid.

3. Privacy Concerns with MOOCs

In this section, we analyze privacy concerns with MOOCs. We focus on three aspects: potential exploitation of student data in business models, unmonitored research use of student data, and limited transparency and data control offered by MOOC providers.

3.1 Business Models

To understand when data collection and usage is helpful and when it's harmful, it is important to consider business models currently pursued by MOOCs, such as certification and job matching, as well as models that might be used in the future, such as targeted advertising. Privacy concerns stem from a mismatch between data use and students' expectations or interests, both of which vary based on the type of business model employed by a MOOC provider. For each business model, we analyze aspects which infringe—or might in the future infringe—upon student privacy.

Certification model

The most common and benign use of student data occurs with certification, the same model followed by traditional schools. By certifying the successful completion of an educational program, a degree signals to the job market the skills acquired by a student.

MOOC providers are partnering with increasingly more higher education institutions as they explore accreditation possibilities, aiming to monetize their MOOCs. Currently, all three major MOOC providers—Coursera, edX and Udacity—offer certificates upon completion for some of their course offerings. While the course material is usually free, the certificates require students to pay a small fee that is currently under $200.[40]

When MOOCs rely on certification as their business model, collection and use of data needed to service the students and deliver the certificates is an expected part of the process. Such data collection—including that of directory information (such as name, address, and date of birth), assignment grades, and exam scores—ultimately benefits the students by enabling them to receive valuable certificates and future degrees. In the case of MIT's MicroMasters program, for instance, certificates from edX can be used to receive transfer credits at various universities, which necessitates disclosure of student information.[41] Students are aware of this and benefit from it. As a result, use of student data solely for certification purposes is unlikely to infringe on their privacy. However, privacy violations can still occur if the student data is discreetly shared with transfer institutions during or after the certification process, or with other third parties without a student's knowledge.

[40] MyLeanMBA. "Breaking down the top 3 MOOC platforms: Coursera, Udacity & edX."
[41] Straumsheim, Carl, "MIT Deems MicroMasters a Success."

Job Matching Model

The job matching model is also currently in use, and is more likely to infringe on student privacy. This business model is essentially a recruiting service, matching users with companies looking to hire students—often with students' permission. By drawing an accurate and multidimensional profile of students, MOOC platforms help employers identify recruiting prospects. Platforms also use data mining to better advise students and to help them present their skills in a more convincing way, thereby facilitating job placement.[42]

Udacity has adopted this model with what was originally called its Talent Source service. As Udacity puts it: "Each Udacity student has a unique profile that can be highly optimized to represent their best work and showcase their most valuable skills. Collectively, these student profiles are organized into an exclusive database."[43]

The downside of such a model is that it begins to blur the line between a student's academic record and a profile for outside opportunities. Without students having transparency and control over how all their data is disclosed, MOOCs could exploit students' information in ways that negatively hurt them in the job market for years to come. Furthermore, because providers like Udacity allow students to sign up with their Facebook or Google account, providers can obtain personal and arguably unrelated information that might jeopardize a student's employability.

Targeted Advertising Model

The business model that has the most potential to raise privacy concerns is targeted advertising. Used by companies across the internet, this model consists of mining as much data as possible about a user and customizing the displayed advertisements accordingly. For example, Facebook targets ads based on dozens of personal data points, including ethnicity, net worth, and relationship status.[44] Unlike with job matching, targeted advertising usually occurs without explicit permission from the user.

While MOOC providers are not currently utilizing targeted advertising, there is little in the way of law to stop them. If nothing changes in the law, it is likely that this business model will be employed in the future. Many other websites, including educational ones, already use targeted advertising. Content-based blogs and how-to websites frequently monetize through hosting advertisements. If regulation of MOOCs and digital student privacy remains ambiguous, it is possible that for-profit companies like Coursera and Udacity will look to introduce new revenue sources through targeted advertising.

---

[42] Belleflamme, Paul and Julien Jacqmin, "An Economic Appraisal of MOOC Platforms: Business Models and Impacts on Higher Education," 163.
[43] Diaz-Hernandez, Ana, "Announcing Talent Source: Connecting Students & Employers With Our New Candidate Sourcing Tool."
[44] Dewey, Caitlin, "98 personal data points that Facebook uses to target ads to you."

The privacy risks that come with targeted advertising on MOOCs are significant. Students' data would essentially be turned into a commodity sold to advertisers, with no direct educational benefit to the students—except perhaps through keeping the MOOC provider in business. Given the big data algorithms used in targeted advertising and lack of transparency in the arena, such practices would likely not be conveyed to students in a clear way. Furthermore, if a student wanted to learn about a topic that no other MOOCs but the one with targeted advertising offered, the student would be compelled to participate in her own data exploitation.

Additionally, advertisements often target users based on sensitive data stored in profiles. Students might therefore have to deal with unexpected or unwanted advertisements that consistently remind them of their disabilities, slow learning rates, mental health conditions, and disciplinary records. Such data uses could affect a student's self-esteem and lead to them dropping out as a result of feeling spotlighted or stigmatized.[45] By leaving MOOCs altogether, students would be forfeiting the benefits of self-paced and personalized learning.

Moreover, targeted ads that are non-educational, such as those selling snack foods or video games, distract students from focusing on learning. Such effects go against society's interest of successfully educating students.[46]

Multiple stakeholders have expressed their concerns with targeted advertising on educational tools, including various states' attorneys general. An example is when lawsuits were filed because of how Google Apps for Education was mining student data by scanning student emails and developing profiles for targeted advertising.[47] MOOCs could follow a similar route, considering how much data they gather for any one student. Whatever the scenario, commercial interests can easily diverge from educational interests with targeted advertising.

3.2 Research Uses

Research can be carried out either by in-house researchers within a MOOC provider or by academics at an outside research university, and both raise some privacy concerns.

In-house researchers and data analytics pose the most significant privacy concerns because no existing law regulates their behavior. Commercial MOOCs have no standard equivalent to a university's institutional review board (IRB). Similarly, commercial MOOCs have no clearly articulated standards governing employee data sensitivity training, the way research should be conducted, and which practices are prohibited or allowed. Without such

[45] Foxman, Maxwell, Mateescu Alexandra, and Bulger Monica. "Advertising in Schools," 2.
[46] DeNisco, Alison. "Protecting student data in the age of marketing and advertising."
[47] Bowden, Tracy, "Google Apps for Education mining data to develop targeted ads, experts warn."

protocols, it may be possible for any employee within a MOOC provider to access a student's personal information and use it in inappropriate ways. For these reasons, it is problematic from the perspective of student privacy for every employee within a large MOOC provider to be able to access sensitive student data.

On the other hand, academic researchers with access to MOOC data pose fewer risks because they are bound by departmental approval, peer review, and approval by Committee on the Use of Humans as Experimental Subjects (COUHES) or an analogous IRB.[48]

Some MOOC providers go further than others in ensuring academic researchers follow these rules and protect student privacy in practice. For instance, edX and partner institutions are at the vanguard of MOOC privacy best practices. EdX devotes entire chapters and sections of its research guide to responsibilities of data teams, including familiarity with data privacy and best practices for security, student data use, and storage.[49] To access data from edX, partner institutions follow a research guide that is over 200 pages.[50] The guide includes detailed instructions for interacting with the "edX Data Package," which includes an event log (or clickstream), exports of course content including databases and forums, and email opt-in data.[51] Students are also made broadly aware of how their data is used. In edX's Privacy Policy, there are guarantees that "publications or public disclosures will not include Personal Information" and that "third party recipients are required to handle the Personal Information in a confidential manner and to maintain adequate security."[52]

However, these procedures and the transparency around them are still largely dependent on a MOOC provider's chosen policies. Commercial MOOC providers like Coursera and Udacity are far less transparent than edX with their data practices.[53] Moreover, even EdX, which explicitly complies with FERPA,[54] may not extend the same privacy protections to users who fall outside of FERPA's scope.

Both in-house and academic research with MOOCs also open up opportunities for student harms that have been long-standing concerns in the education sector: predictive sorting, filter bubbles, and discrimination.[55] With massive amounts of detailed student data, researchers have the ability to categorize students in many ways, including those that could be unfairly discriminating. For example, students from a predominantly poor demographic might be erroneously categorized as slow learners, while the true explanation may be the lack of resources and study materials. Such discriminatory effects infringe on students' privacy since the students are not aware that their data is being used to classify them in

[48] MIT Institutional Research, "MITx Data Request Checklist."
[49] edX, "EdX Research Guide," 6-17.
[50] Ibid.
[51] Ibid, 3.
[52] edX, "Privacy Policy."
[53] The authors could not find any publicly-available data-handling guidance even remotely analogous to that published by edX.
[54] edX, "Privacy Policy."
[55] Jones, Meg Leta and Lucas Regner, "Users or Students? Privacy in University MOOCs," 1493.

such a damaging way.[56] Even if they were aware, they might then have to modify their behavior unnaturally or otherwise make up for their lack of control over the data that MOOCs collect and use.

3.3 Limited Data Transparency and Control

Beyond the privacy risks arising from both the business models and research applications, MOOC providers, intentionally or unintentionally, jeopardize students' privacy by limiting student access to and control over personal data. These basic privacy concerns extend throughout all MOOC and edtech services that exist today.

Parents and students currently have no guaranteed access to their data on MOOCs. It is not generally possible for a student to inspect stored data, to report data inaccuracies, or to request deletion in case they want to leave the platform and leave no data behind.

Even edX lacks transparency and control provisions in its policies. This is problematic because it is the most transparent MOOC provider in terms of data practices and procedures. Yet even if a student reads the privacy policy and research guides, she still may not know to what extent her data is collected, to whom it is distributed, and to what end it is used or analyzed.

Furthermore, the inconsistency in data policies across MOOCs—when they exist—leave students confused and uncertain of how exactly their privacy is protected. For instance, Coursera provides a limited data dashboard that lists only a portion of the actual user data collected, while edX on the other hand provides no such dashboard. Instead, an edX user may email a provided address to request access to the information maintained.[57] Neither solution is especially desirable in terms of upholding student privacy rights or shedding light on opaque student data practices.

3.4 Case Study: inBloom

The former company inBloom is a prime example of students' lack of data transparency and control online, as well as the ensuing consequences.

Founded in 2011, inBloom was funded by the Gates Foundation to run a student data collection project to aggregate a wide range of student data from various states and districts. The goal was to collect and integrate student attendance, assessment, disciplinary and other records from disparate school-district databases, put the information in cloud storage, and release it to authorized web services that could help teachers track student progress.[58]

[56] Jones, Meg Leta and Lucas Regner, "Users or Students? Privacy in University MOOCs," 1493..
[57] Ibid., 1490.
[58] Monica Bulger, Patrick McCormick, and Mikaela Pitcan, "The Legacy of inBloom," 6.

Parents and privacy advocates questioned the motive behind the organization holding a vast amount of student data that included discipline records. They were also concerned by how easy inBloom had made it for third parties to access student data without the students and parents' consent.[59] In other words, inBloom had given students and parents little control over how their data was used.

Public criticism grew and the company was forced to cease operations after only fifteen months. InBloom's shutdown demonstrates parents' concerns with how much student data was being collected without their knowledge or consent. This information was especially sensitive; in addition to disciplinary records, it included income level, disabilities and Individual Education Plans (IEPs), mental health records, and medical history.[60] Parents also discovered that there were other initiatives supported by the Federal government and the Gates Foundation that were pushing for student data sharing with for-profit data-mining vendors and other third parties, also without parent notification or consent.[61]

Although inBloom focused on K–12 schools, the outcome generalizes to MOOCs and illustrates how crucial it is that adequate laws are enacted to address student data privacy concerns. MOOCs mine massive amounts of student data similar to how inBloom did, and without proper data transparency and control, MOOCs might create similar problems and face comparable backlash.

4. Current Approaches to Online Privacy

We now consider current approaches that try to counter the privacy problems from Section 3, without eroding the innovation and benefits from Section 2. We take special note of these approaches' shortcomings.

4.1 FERPA Amendments and Guidance

Although FERPA was signed into law in 1974, it has since been reinterpreted and amended numerous times.[62] This has allowed for some, although limited, advancements in extending student privacy rights online.

FERPA was most recently amended in 2008 and 2011.[63] 2008 marked the greatest change in FERPA's scope as it relates to technology vendors working with schools. While most disclosures of student information are prohibited, a key exception in the original FERPA law is the "school official" clause. While this had originally been intended to encompass

[59] Monica Bulger, Patrick McCormick, and Mikaela Pitcan, "The Legacy of inBloom," 6.
[60] Strauss, Valerie, "The astonishing amount of data being collected about your children."
[61] Haimson, Leonie and Rachael Stickland, "About Us."
[62] Electronic Privacy Information Center (EPIC), "Family Educational Rights and Privacy Act (FERPA)."
[63] Ibid.

people like teachers and cafeteria employees, the 2008 amendment expanded it to explicitly include vendors like technology companies.[64] In particular, it allowed schools to share data without a parent's permission to a "contractor, consultant, volunteer, or other party to whom an agency or institution has outsourced institutional services or functions may be considered a school official,"[65] as long as the party is under the school's control and using the data for educational purposes. Technology vendors now fell under this new definition of "school officials," allowing for schools to outsource services and data to these companies.[66]

In addition to amendments, the Department of Education has offered regulatory guidance on how FERPA applies to online companies. It founded the Privacy Technical Assistance Center (PTAC) in 2010 to act as a "'one-stop' resource for education stakeholders to learn about data privacy, confidentiality, and security practices related to student-level data systems and other uses of student data."[67] The Center is run by the Chief Privacy Officer of the Department of Education.[68] Its primary guidance for online educational services came in 2014 in the form of a report titled "Protecting Student Privacy While Using Online Educational Services: Requirements and Best Practices."[69] [70]

However, despite these updates to FERPA and its interpretation, the law continues to be severely insufficient for addressing modern student privacy concerns. Most notably, the law only pertains to schools receiving federal funds and the vendors they interface with. It does not cover online websites that don't work directly with schools, including most MOOCs. The outdatedness of FERPA can also be seen clearly in the "school officials" exception. Because there is no concept of technology vendor in FERPA, the law instead groups teachers and technology companies together as school officials, granting them the same rights when working with student data. This results in there being very little guidance for vendors working with sensitive student data.[71]

Even if FERPA did apply to MOOCs appropriately, the law's enforceability remains questionable. Parents do not have the right to sue[72], and as law professor and privacy expert Daniel Solove writes, the only sanction that exists under FERPA is "a sanction so implausible it has never been imposed in the 35+ year history of the law. That sanction is a withdrawal of all federal funds. It will never happen."[73] Furthermore, technology vendors are one step removed from even that sanction.[74] As such, it is no surprise that the guidance from PTAC, while helpful, carries little weight. PTAC even acknowledges itself that

[64] Polonetsky, Jules and Omer Tene, "Who is Reading Whom Now: Privacy in Education from Books to MOOCs," 963-964.
[65] LII, "34 CFR 99.31 - Under what conditions is prior consent not required to disclose information?" (a)(1)(i)(B).
[66] Polonetsky, Jules and Omer Tene, "Who is Reading Whom Now: Privacy in Education from Books to MOOCs," 964.
[67] U.S. Department of Education (ED), "About Us."
[68] Ibid.
[69] PTAC, "Protecting Student Privacy While Using Online Educational Services: Requirements and Best Practices."
[70] Fitzpatrick, Kaleigh C., "Student Data at Risk: A Multi-Tiered Approach for Massachusetts to Mitigate Privacy Risks While Utilizing Innovative Education Technology in Schools," 313.
[71] Polonetsky, Jules and Omer Tene, "Who is Reading Whom Now: Privacy in Education from Books to MOOCs," 967.
[72] Fitzpatrick, Kaleigh C., "Student Data at Risk: A Multi-Tiered Approach for Massachusetts to Mitigate Privacy Risks While Utilizing Innovative Education Technology in Schools," 306.
[73] Solove, Daniel, "FERPA and the Cloud: What FERPA Can Learn from HIPAA."
[74] Polonetsky, Jules and Omer Tene, "Who is Reading Whom Now: Privacy in Education from Books to MOOCs," 967.

"FERPA represents a minimum set of requirements to follow."[75] Clearly, FERPA is currently serving more as guidelines than enforceable law for providers of new technology.

4.2 California's SOPIPA

In 2014, the state of California recognized the increasing number of privacy concerns with private MOOCs and sought to regulate them through the Student Online Personal Information Protection Act (SOPIPA).[76] The landmark statute is meant to fill the holes in federal laws like FERPA by preventing student data from being sold and by targeting companies and not just schools.[77] The law passed in California, which is a leader in privacy law, but applies to any edtech company serving California students. Thus, its reach extends well beyond California.[78]

SOPIPA went into effect in January, 2016 and set a new standard for how edtech websites could be regulated. The law limits "commercial advertising, marketing and profiling by operators of websites or providers of Internet services or mobile applications,"[79] pertaining to any "operator of an Internet Website, online service, online application, or mobile application with actual knowledge that the site, service, or application is used primarily for K–12 school purposes and was designed and marketed for K–12 school purposes."[80] SOPIPA also prevents edtech companies from selling student data or amassing student profiles for commercial purposes.[81] [82] It additionally gives schools the right to delete any data on an edtech company's website that is under the school's control.[83] Furthermore, although it does not mandate it, the law allows edtech companies to offer students the ability to download or export their data—giving students control over their information. All of these regulations were unprecedented until SOPIPA came along.[84]

SOPIPA was meant to restrict commercial uses of student data, but to continue to allow research uses. In doing so, California lawmakers hoped to assuage the concerns of parents. A 2014 study by Benenson Strategy Group revealed that 86% of adults believe that "oversight is necessary to ensure [children's] private information is not exploited for commercial purposes and stays out of the hands of the wrong people."[85] Lawmakers drew a clear distinction between commercial and research uses of student data. While most disclosures of personally identifiable information are prohibited, edtech companies are still

[75] PTAC, "Protecting Student Privacy While Using Online Educational Services: Requirements and Best Practices," 5.
[76] California Senate, "Senate Bill No. 1177 Chapter 839."
[77] Harris, Kamala D., "Ready for School," 8.
[78] Peterson, Dylan, "Edtech and Student Privacy: California Law as a Model," 973.
[79] Halpert, Jim and Michelle Anderson, "State Privacy and Security Developments–Looking Back and Looking Ahead."
[80] California Senate, "Senate Bill No. 1177 Chapter 839," 22584(a).
[81] Peterson, Dylan, "Edtech and Student Privacy: California Law as a Model," 973.
[82] EPIC, "State Student Privacy Policy."
[83] Ibid.
[84] Future of Privacy Forum, "FPF Guide to Protecting Student Data Under SOPIPA," 21.
[85] Common Sense Media, "Student Privacy Survey."

able to share such information with researchers, as long as it's under the "direction of a school, school district, or state department of education."[86]

While SOPIPA does not include any explicit enforcement provisions, it is meant to be enforced through California's broad Unfair Competition Law (UCL), which covers most unlawful business practices.[87] The law allows California's Attorney General, district attorneys, and certain city attorneys to file an action for unfair competition, which could result in fines. It also permits individuals to seek recourse in some cases.[88] However, because SOPIPA was only enacted recently, it remains unclear how courts will react to such cases.[89]

Despite representing a significant step forward in protecting student privacy, SOPIPA suffers from numerous drawbacks. Perhaps most notably, it fails to protect the rights of students outside K–12 and other learners who utilize educational websites. Instead, it restricts its scope by continually referring to "K–12 student[s]."[90] Although California has since passed the Early Learning Personal Information Protection Act to apply the same protections to students in preschool and prekindergarten, they have not yet extended the protections to students who are beyond twelfth grade, such as university students and adult learners.[91]

Additionally, SOPIPA's scope is limited in that it does not do much to provide students with transparency around data that is collected about themselves.[92] Although openness is one of the main principles in the Organisation for Economic Co-operation and Development's (OECD) Privacy Framework, the law does not require edtech companies to reveal their data collection or use practices.[93] [94] SOPIPA also fails to provide students and parents with proper control over their data. Only schools, and not students, are able to delete data stored on educational websites under the law.[95]

Finally, the enforcement of SOPIPA provisions is likely too weak and limited given the ambiguities in the law. When imposing security requirements, for instance, SOPIPA states that companies must have "reasonable security procedures and practices appropriate to the nature of the covered information."[96] However, it does not specify what "reasonable" or "appropriate" means. And although it is too soon to know how enforcement will fare under California's UCL, the absence of a definition for targeted advertising in SOPIPA will likely

[86] California Senate, "Senate Bill No. 1177 Chapter 839," 22584(e)(2).
[87] Future of Privacy Forum, "FPF Guide to Protecting Student Data Under SOPIPA," 23.
[88] Peterson, Dylan, "Edtech and Student Privacy: California Law as a Model," 975.
[89] Future of Privacy Forum, "FPF Guide to Protecting Student Data Under SOPIPA," 24.
[90] California Senate, "Senate Bill No. 1177 Chapter 839," 22584(a).
[91] Harris, Kamala D., "Ready for School," 6.
[92] Peterson, Dylan, "Edtech and Student Privacy: California Law as a Model," 993.
[93] OECD, "The OECD Privacy Framework," 15.
[94] Peterson, Dylan, "Edtech and Student Privacy: California Law as a Model," 994.
[95] Ibid., 993.
[96] California Senate, "Senate Bill No. 1177 Chapter 839," 22584(a).

make meeting the thresholds of proof in the UCL nearly impossible.[97] The law may also have loopholes that could lead to the commercial use of student data under the guise of educational purposes.[98] Without proper transparency provisions in the law, these commercial uses could easily go undetected. It is also possible that the UCL is the wrong enforcement method altogether. Because SOPIPA does not explicitly include a right of action against edtech companies that violate student privacy rights, there is no unambiguous way for students to protect their rights if California does not take action itself.[99]

Between its limited K–12 scope, lack of transparency and control provisions, and ambiguities in definitions, SOPIPA falls short of its goal of protecting student privacy online.

4.3 Student Privacy Pledge

Soon after SOPIPA passed, a group of edtech companies joined a voluntary Student Privacy Pledge, introduced by The Future of Privacy Forum and the Software & Information Industry Association.[100] They agreed to not sell information on K–12 students and to not utilize targeted advertising on their educational services. In other words, they took the SOPIPA principles and brought them to a national level. Some speculate that these companies took it upon themselves to do this in order to "fend off tighter regulation" by "plug[ging] some of the loopholes in federal privacy law."[101]

The pledge does improve upon the privacy rights included in SOPIPA, but only in a few ways. For instance, it focuses more on transparency, requiring signatories to "[d]isclose clearly in contracts or privacy policies, including in a manner easy for parents to understand, what types of student personal information we collect, if any, and the purposes for which the information we maintain is used or shared with third parties."[102]

Since it was released, the pledge been supported by numerous parties. This includes members of Congress such as Representatives Jared Polis (D-CO) and Luke Messer (R-IN), as well as President Barack Obama.[103] [104] It also includes signatories like Khan Academy, Gradescope, Apple, and Google, as well as over 300 other companies.[105] [106] It does not, however, include edX, Coursera, and Udacity.[107]

[97] Future of Privacy Forum, "FPF Guide to Protecting Student Data Under SOPIPA," 24.
[98] Peterson, Dylan, "Edtech and Student Privacy: California Law as a Model," 992.
[99] EPIC, "State Student Privacy Policy."
[100] Harris, Kamala D., "Ready for School," 3.
[101] Peterson, Dylan, "Edtech and Student Privacy: California Law as a Model," 975-976.
[102] Student Privacy Pledge, "Privacy Pledge."
[103] Student Privacy Pledge, "Student Privacy Pledge Reaches Milestone of 300 Signatories."
[104] Peterson, Dylan, "Edtech and Student Privacy: California Law as a Model," 976.
[105] Student Privacy Pledge, "Signatories."
[106] Alim, Frida et al., "Spying on Students," 24.
[107] Student Privacy Pledge, "Signatories."

Therein lies the downside: the pledge, while enforceable by the Federal Trade Commission (FTC) once agreed to, is voluntary to begin with.[108] As such, none of the Big Three MOOC providers have agreed to it.

Another downside is that like SOPIPA, the scope is limited. The pledge is once again geared towards only K–12 edtech companies. It also does not include any guidelines for providing students with control over their data or allowing them to delete it.[109]

The Privacy Pledge also suffers from a number of potential loopholes that could arise from its vague definitions and principles, similar to those in SOPIPA.[110] Targeted advertising is once again not defined, and is instead described as "behavioral targeting of advertisements to students."[111] Security requirements include phrases like "reasonably designed."[112]

In general, industry self-regulation has severe limitations in that enforcement is practically non-existent. With the Privacy Pledge, although violations could potentially result in FTC action for "unfair and deceptive business practices,"[113] the evidence shows that this is not happening. The Electronic Frontier Foundation, for instance, filed an FTC complaint in 2015 against Google for violating provisions in the pledge, but at the time of writing there have been no public indications of actions taken by the FTC.[114] Thus, even in cases where companies choose to join the Privacy Pledge and the pledge's definitions are properly specified, any violations are unlikely to be enforced.

4.4 State Laws

California's SOPIPA, passed in 2014, also initiated a renaissance in student privacy laws across the country. Since 2013, over 400 bills have been introduced in state legislatures to address student privacy concerns. As of September 2016, 73 of those bills had been passed into law by 36 states. Many of these laws focus on online privacy and are modeled after SOPIPA. In 2016 alone, nearly half of the introduced student privacy bills concerned third party providers, and one third of them contained provisions based on SOPIPA's.[115]

While many of the provisions in these other states' laws are the same as those found in SOPIPA, some evolution has occurred. In 2016, 23 of the laws defined targeted advertising, whereas SOPIPA did not.[116] As one of the criticisms of SOPIPA is its ambiguities in definitions, these laws may be more easily enforceable. Many of the laws adopted a definition similar to that of 2015 legislation introduced in Congress: targeted advertising is

[108] Student Privacy Pledge, "FAQs."
[109] Student Privacy Pledge, "Privacy Pledge."
[110] Alim, Frida et al., "Spying on Students," 24.
[111] Student Privacy Pledge, "Privacy Pledge."
[112] Ibid.
[113] Alim, Frida et al., "Spying on Students," 25.
[114] Ibid.
[115] Data Quality Campaign, "Student Data Privacy Legislation: A Summary of 2016 State Legislation," 1-3.
[116] Ibid., 3.

"presenting advertisements to a student or the student's parent, where the advertisements are selected based on information obtained or inferred from the student's online behavior or use of online applications or mobile applications or from covered information about the student maintained by the operator of a school service."[117] [118] Some states have begun to include more explicit enforcement provisions in their laws. Idaho, for example, imposes penalties for violating its student privacy law.[119] Additionally, two states, Colorado and Connecticut, have begun to differentiate between edtech companies that have formal contracts with schools and those that don't.[120]

However, these laws still suffer from the same general problems as SOPIPA. The Data Quality Campaign notes that "[s]tates will continue to iterate, but clearly SOPIPA has become the standard."[121] Thus, most of these laws inherit the same limited scope that excludes MOOCs, the lack of transparency and control provisions, and the weak enforcement mechanisms.

4.5 President Obama's Proposed Legislation

While no federal law akin to SOPIPA has passed, there have been several proposals. The most notable is President Barack Obama's legislative proposal that he unveiled in January 2015.[122] During his speech, he highlighted the importance of student privacy and said that "data collected on students in the classroom should only be used for educational purposes—to teach our children, not to market to our children."[123]

President Obama acknowledged both SOPIPA and Student Privacy Pledge as models for his proposal. The bill was introduced in the House as the Student Digital Privacy and Parental Rights Act of 2015, a bipartisan effort by Representatives Jared Polis (D-CO) and Luke Messer (R-IN).[124] However, it failed to pass in the Republican-held Congress.[125]

The bill improved upon SOPIPA just as state bills after it did by including a definition of targeted advertising.[126] It also gave parents and not just schools control of their children's data, including the ability to delete data that is not needed by schools anymore.[127] The provisions in the bill were supposed to be enforced by the FTC, with the help of the Department of Education when schools were involved.[128]

---

[117] Data Quality Campaign, "Student Data Privacy Legislation: A Summary of 2016 State Legislation," 3.
[118] Messer, Luke. "Student Digital Privacy and Parental Rights Act of 2015," 2(a)(13)(A).
[119] Parent Coalition for Student Privacy, "State Student Privacy Laws."
[120] Data Quality Campaign, "Student Data Privacy Legislation: A Summary of 2016 State Legislation," 5.
[121] Ibid.
[122] Harris, Kamala D., "Ready for School," 7.
[123] The White House Office of the Press Secretary, "Remarks by the President at the Federal Trade Commission."
[124] Singer, Natasha, "Legislators Introduce Student Digital Privacy Bill."
[125] Congress.gov, "H.R.2092 - Student Digital Privacy and Parental Rights Act of 2015."
[126] Congress.gov, "H.R. 2092," 2(a)(13).
[127] Singer, Natasha, "Legislators Introduce Student Digital Privacy Bill."
[128] Roscorla, Tanya, "3 Student Data Privacy Bills That Congress Could Act On."

However, even if the bill were to have passed, it suffers from many of the same limitations as the state laws. It would have only applied to K–12 students.[129] While it would have offered control and transparency to schools, it would not have afforded the same rights to students themselves.[130] Additionally, some of the provisions of SOPIPA were excluded, including amassing profiles of students for commercial purposes. As a result, the proposal faced criticism from many privacy advocates.[131]

5. Legislative Recommendation

5.1 Proposing a New Federal Law

Given the limitations in the current legislation and approaches to student privacy, we propose the introduction of a new federal law based on California's SOPIPA, but adapted for the federal government and applicable to all educational websites, including K–12 edtech companies and MOOCs. Such a Student Internet Privacy Act would protect all students' privacy online in the United States.

5.2 Principles Underlying Legislation

A standard education privacy framework is necessary to guide the formulation of future education privacy laws and policies, both in the United States and abroad. Such a framework could serve a role similar to the privacy principles published by the OECD.[132] The framework could also guide the creation of enforceable pledges like the Student Privacy Pledge.

It is important to note that we are not the first to propose such a framework. For instance, the Parent Coalition for Student Privacy, which was founded in 2014 to advocate for student privacy following inBloom's student data controversies, has laid out its own five principles.[133] Therefore, drawing upon previous work by many others, we summarize four key education privacy principles that cover the basic rights of students and other standards recognized by society. These principles serve as the basis for our legislative recommendations.

Principle 1: Transparency

The Transparency principle is a prerequisite for all other privacy principles. Transparency in educational privacy aligns with the Openness principle in the OECD's guidelines for the

[129] Congress.gov, "H.R. 2092," 2(a)(6).
[130] Ibid., 2(b) especially 2(b)(3).
[131] Herold, Benjamin, "Draft of President Obama's Student-Data-Privacy Bill Raises Questions."
[132] OECD, "The OECD Privacy Framework," 14-15.
[133] Parent Coalition for Student Privacy, "Five Principles to Protect Student Data Privacy."

protection of privacy,[134] and the transparency right as stipulated in the consumer privacy bill of rights that was released by Obama's administration.[135] It requires that users of a service or product, such as a MOOC, be cognizant of data collected about them and how it is used. It implies that a privacy violation has occurred if a service provider misuses personal data, uses data for unauthorized purposes, or shares it with an unauthorized third party. In 2014, the Electronic Privacy Information Center (EPIC) proposed a student bill of rights in which they placed emphasis on transparency from schools and online education service providers with regards to what student data is collected, stored, and used.[136] Such a Transparency principle can reduce privacy risks, such as unauthorized data collection. Serious misuses like those of inBloom might be prevented if students knew more about a website's data collection.[137]

Principle 2: Control of Data

A student should have the ability to view data collected in a usable format, to verify its accuracy, to request for modifications in case of inaccuracies, and to request for its deletion. This follows from OECD's Individual Participation principle[138] as well as Individual Control from the FTC's FIPPs.[139] Data Control is essential in this age where technology has made it possible for data controllers, as defined in OECD's guidelines,[140] to process and use massive amounts of data in numerous ways. Students deserve to be in control of their information as this could help mitigate risks such as inaccurate representation of students or disclosure of sensitive matters like learning disabilities.[141]

This principle was also one of the primary drivers of the FERPA legislation. Senator Buckley proposed FERPA to Congress in 1974 at a time when both parents and students had less access to student education records and furthermore had no guaranteed right that records could be updated when there were discrepancies. This was alarming to many parents and organizations, and it laid a foundation for education privacy legislation and future principles.[142]

The FTC's FIPPs specifically note that students and parents "have an interest in assuring that whatever information web sites collect from children or have otherwise obtained about their children is accurate,"[143] and that it "is particularly important in contexts that involve decisions that impact on the child or family, such as educational or health decisions."[144]

---

[134] OECD, "The OECD Privacy Framework," 15.
[135] The White House Office of the Press Secretary, "We Can't Wait: Obama Administration Unveils Blueprint for a 'Privacy Bill of Rights' to Protect Consumers Online."
[136] EPIC, "Student Privacy Bill of Rights."
[137] Peterson, Dylan, "Edtech and Student Privacy: California Law as a Model," 994.
[138] OECD, "The OECD Privacy Framework," 15.
[139] Federal Trade Commission, "Priavcy Online: A Report to Congress," 7-11.
[140] OECD, "The OECD Privacy Framework," 13.
[141] Ibid., 13 (especially Scope of Guidelines 2).
[142] Jones, Meg Leta and Lucas Regner, "Users or Students? Privacy in University MOOCs," 1479.
[143] Federal Trade Commission, "Priavcy Online: A Report to Congress," 14.
[144] Ibid.

With control of data, students can ensure their data is kept accurately. Furthermore, students have the ability to delete their data if they feel their privacy is at risk.

Principle 3: Focused Collection

Focused Collection refers to a practice where education providers only collect the amount of data they need while clearly disclosing the extent of data collection and use of the data. Focused Collection draws upon the OECD's Collection Limitation Principle and the Choice/Consent principle from the FTC's FIPPs.[145] [146] Both principles, though from different sources, acknowledge the exponentiated risks of large amounts of student data collection and thus propose requiring providers to be more purposeful when collecting any data from students. They also related to the idea that data should not be kept longer than necessary.

There have been multiple news articles lamenting the amount of student data that is collected today without the knowledge of students and their parents, due to increased use of technology. Leonie Haimson and Cheri Kiesecker of The Parent Coalition for Student Privacy articulate their worry as follows:

> Remember that ominous threat from your childhood, "This will go down on your permanent record?" Well, your children's permanent record is a whole lot bigger today and it may be permanent. Information about your children's behavior and nearly everything else that a school or state agency knows about them is being tracked, profiled and potentially shared.[147]

Focused Collection would reduce the amount of data available for misuse by a MOOC or by third party providers who could access the data through purchasing it, whether legal or not. This principle helps assure students that they are not being tracked excessively, and enables them to participate in MOOCs with greater confidence.

Principle 4: Educational Use and Context

The principle of Educational Use and Context implies that student data should not be commercialized for use in marketing or targeted advertising. This principle goes hand in hand with OECD's Purpose Specification and Use Limitation principles. Applied to education, it would require that education providers limit their uses of student data to education research and learning analytics aimed at enhancing teaching methods.[148] OECD's Focused Collection principle also applies, as data that's not necessary for education purposes becomes unnecessary and thus should not be collected.

EPIC has consistently fought against student privacy threats in areas where data is beyond educational purposes. The leading student privacy advocate sued the Department of

[145] OECD, "The OECD Privacy Framework," 14.
[146] Federal Trade Commission, "Priavcy Online: A Report to Congress," 8.
[147] Strauss, Valerie, "The astonishing amount of data being collected about your children."
[148] OECD, "The OECD Privacy Framework," 14.

Education in 2015 for changing FERPA to include regulations that would allow third parties, including private companies and foundations promoting school reform, to gain access to student data.[149] President Obama is another proponent who has voiced his support for this principle. In his 2015 speech on proposed new actions to protect Americans' privacy he addresses this issue:

> "Today, we're proposing the Student Digital Privacy Act. We're saying that data collected on students in the classroom should only be used for educational purposes—to teach our children, not to market to our children. We want to prevent companies from selling student data to third parties for purposes other than education. We want to prevent any kind of profiling that outs certain students at a disadvantage as they go through school."[150]

Students require protection against uses of data that could stigmatize them, expose their disabilities or inabilities, or prevent them from getting jobs. Allowing only educational use of student data and in a research context is an important step towards meeting students' needs and protecting their privacy.

5.3 Definitions

In order to introduce a new federal law in the United States that will realize the above four principles, the appropriate definitions must be in place. While SOPIPA broadens the scope of regulated actors to include K–12 edtech websites and not just schools, we suggest broadening the scope even further to protect every learner in the online realm.

Operator

In order to encompass both edtech companies serving K–12 students and MOOCs serving students of all ages, we recommend expanding the SOPIPA definition of "operator" from:

> *The operator of an Internet Website, online service, online application, or mobile application with actual knowledge that the site, service, or application is used primarily for K–12 school purposes and was designed and marketed for K–12 school purposes.*

to:

> *Either a K–12 edtech operator or MOOC operator, given the following definitions*
> *1) "K–12 edtech operator": the operator of an Internet Web site, online service, online application, or mobile application with actual knowledge that the site, service, or application is used primarily for K–12 school purposes and was designed and*

[149] Strauss, Valerie, "Lawsuit charges Ed Department with violating student privacy rights."
[150] Hudson, David, "The President Announces New Actions to Protect Americans' Privacy and Identity."

marketed for K–12 school purposes
2) "MOOC operator": the operator of an Internet Web site, online service, online application, or mobile application that serves the same primary functions as traditional schools, including courseware and some form of certification.

This definition is meant to include all edtech operators that are encompassed under SOPIPA, as well as MOOCs like the Big Three and Udemy, which serve some combination of K–12 students, university students, and adults.

Student

Throughout the bill, "student" should refer to any user of the operator's online service. It is important to note that while SOPIPA only protects K–12 students on edtech operators like Khan Academy, this law would cover all users.

In establishing this scope, we ensure that our four aforementioned privacy principles are extended to all learners who access a website with the expectation of receiving an education—not just a subset of them. As established in Section 1.4, educational privacy rights are based on violations that could hurt people in educational settings. They therefore don't need to be restricted to only K–12 students.

Finally, if a student is not yet a legal adult, the parents of the student should be afforded the equivalent rights.

Personally Identifiable Information

We recommend maintaining the current definition of personally identifiable information (PII), or "covered information," used by SOPIPA. This definition encompasses the many types of big data that can be collected on students, such as search history and geolocation, which were discussed in Section 3.1 when considering privacy risks that arise with MOOC business models.

Personally identifiable information or materials, in any media or format that meets any of the following:

(1) Is created or provided by a student, or the student's parent or legal guardian, to an operator in the course of the student's, parent's, or legal guardian's use of the operator's site, service, or application for [educational] purposes.

(2) Is created or provided by an employee or agent of the K–12 school, school district, local education agency, or county office of education, to an operator.

(3) Is gathered by an operator through the operation of a site, service, or application described in subdivision (a) and is descriptive of a student or otherwise identifies a

student, including, but not limited to, information in the student's educational record or email, first and last name, home address, telephone number, email address, or other information that allows physical or online contact, discipline records, test results, special education data, juvenile dependency records, grades, evaluations, criminal records, medical records, health records, social security number, biometric information, disabilities, socioeconomic information, food purchases, political affiliations, religious information, text messages, documents, student identifiers, search activity, photos, voice recordings, or geolocation information.

Educational Purposes

We propose expanding the "K–12 school purposes" phrase used in SOPIPA, defined as

> *purposes that customarily take place at the direction of the K–12 school, teacher, or school district or aid in the administration of school activities, including, but not limited to, instruction in the classroom or at home, administrative activities, and collaboration between students, school personnel, or parents, or are for the use and benefit of the school,*

to "educational purposes," defined as

> *purposes that customarily take place at the direction of the K–12 or university school, teacher, or school district or aid in the administration of school activities, including, but not limited to, instruction in the classroom or at home, administrative activities, and collaboration between students, school personnel, or parents, or are for the use and benefit of the school,*

which now includes higher education and professional degree programs.

Targeted Advertising

Rather than referring to SOPIPA, which does not include a definition for "targeted" advertising, we adopt a definition from a similar Virginia law written in 2016. This added precision would ensure that our proposed bill could be adequately enforced without introducing a number of loopholes currently found in SOPIPA, as discussed in Section 4.2.

> *"Targeted advertising" means advertising that is presented to a student and selected on the basis of information obtained or inferred over time from such student's online behavior, use of applications, or sharing of student personal information. "Targeted advertising" does not include advertising*
> *(i) that is presented to a student at an online location (a) on the basis of such student's online behavior, use of applications, or sharing of student personal information during his current visit to that online location or (b) in response to that student's request for information or feedback and*

(ii) for which a student's online activities or requests are not retained over time for the purpose of subsequent advertising.[151]

The exception in the definition allows for a MOOC provider to advertise other educational courses on their website, which is a type of personalized recommendation from which students greatly benefit, as mentioned in Section 2.1. Furthermore, a job matching business model can use this exception to continue matching students with employers, as long as the student requests it, which is in line with Principle 2, Control of Data. And in both cases, the targeted advertising exception requires that the data not be retained over time, which satisfies Principle 3, Focused Collection.

5.4 Legislative Clauses

Key Clauses to Adopt From SOPIPA

In order to prevent the exploitative business uses of student data described in Section 3, we adopt the key clauses that make up SOPIPA.

No targeted advertising

(1) "An operator shall not knowingly …
(A) Engage in targeted advertising on the operator's site, service, or application, or
(B) target advertising on any other site, service, or application when the targeting of the advertising is based upon any information, including covered information and persistent unique identifiers, that the operator has acquired because of the use of that operator's site, service, or application" [152]

No student profiles except for educational uses

(2) "An operator shall not knowingly … [u]se information, including persistent unique identifiers, created or gathered by the operator's site, service, or application, to amass a profile about a student except in furtherance of [educational] purposes." [153]

No sale of student information

(3) "An operator shall not knowingly … [s]ell a student's information, including covered information." [154]

Research exemption

Because such a law might otherwise have chilling effects on research and the benefits described in Section 2, we make sure to include the strong research exemption clause from

[151] Virginia General Assembly, "HB 749" 1(A).
[152] California Senate, "Senate Bill No. 1177 Chapter 839," 22584(a)(1).
[153] Ibid., 22584(b)(2).
[154] California Senate, "Senate Bill No. 1177 Chapter 839," 22584(b)(3).

SOPIPA in its entirety. Academic researchers would thus be able to obtain entire datasets, and not just de-identified datasets that might limit their research results. There are few associated privacy risks with this exemption because of the existing laws that already regulate how student data is handled by researchers in university settings.

> *(4) "[A]n operator may disclose covered information of a student … under the following circumstances: … For legitimate research purposes:*
> *(A) as required by state or federal law and subject to the restrictions under applicable state and federal law or*
> *(B) as allowed by state or federal law and under the direction of a school, school district, or state department of education, if no covered information is used for any purpose in furtherance of advertising or to amass a profile on the student for purposes other than [educational] purposes."* [155]

Recommendations for New Clauses

Data Dashboards

Following Principle 1, Transparency, students should have access to a secure data dashboard that outlines the types of data collected and how they're being used. While this may include general descriptions for micro-scale data like clickstream, directory information like age and school should be displayed in full to the student. The purpose of the data usage should be clearly stated, as required by Principle 3, Focused Collection. In line with Principles 2 and 4, Control of Data and Educational Use and Context, operators can request a student's permission to collect and use data for more educational purposes than previously agreed to, as long as the student opts in.

In requiring that every operator offer a data dashboard of some kind, we are lowering the chances of another inBloom case where massive amounts of student data are collected unbeknownst to the students initially.

> *(5) An operator shall provide each student with a "data dashboard."*
> *(a) "Data dashboard" means a secure webpage that describes all types of data that the operator has collected on the student and how each type of data is used.*
> *(b) Data that is directory information or similarly discrete should be disclosed in detail to the student on the secure dashboard, whereas other data types may be described at a higher level.*
> *(c) This section does not limit the ability of an operator to request a student's permission for collecting or sharing new types of data, subject to the following conditions: (a) the student must opt into the new type of data collection or usage on the dashboard, and (b) the data collection or usage must be for*

[155] California Senate, "Senate Bill No. 1177 Chapter 839," 22584(e)(2).

well-defined educational purposes and not targeted advertising or any other non-educational purpose.

Deletion of One's Own Account

Given Principle 2, Control of Data, students should be afforded the option to delete their accounts and all data that a provider has collected on them. The one exception should be data that was previously shared with researchers and can not be easily deleted. This exception mitigates the chilling effect on research and the self-selection bias that might otherwise ensue if research analyses were to neglect the data of deleted users. However, it still upholds the rights of students to delete their data from an operator's website.

(6) An operator shall provide students with a well-defined process for requesting the deletion of their accounts and all associated data collected or produced during the account's lifetime. Within 30 days of the request, the operator shall delete the aforementioned account and data, excluding any research data that was disclosed to researchers under section (4) before the student made the initial request.

Amendment of One's Own Data

Given Principle 2, Control of Data, students should be able to amend inaccurate or misleading data pertaining to themselves. This ensures that students don't have data inappropriately used against them by MOOC providers. Without such a clause, a MOOC provider could potentially make a mistake and send inaccurate information to employers as part of a job matching business model—a scenario which might have negative, long-term effects on a student.

(7) An operator shall provide students with a well-defined process for amending any data associated with them that is inaccurate or misleading. This includes, but is not limited to, data displayed on the operator's data dashboard.

Exporting One's Own Data

Following Principle 2, Control of Data, operators should allow students to download any data that is readily exportable and pertains to the student. This includes all data that operators are already disclosing to researchers. By focusing on data that is readily exportable, we ensure that there is not an undue burden on operators to develop new processes for exporting data.

(8) An operator shall provide students with the option to download, or otherwise export, all data pertaining to them, as long as (a) the data is already being provided to researchers, (b) the data is directory information, or (c) the data is covered information that can readily be exported without serving an undue burden on the operator.

Access Control

Just as human research subjects are protected by the law[156] against privacy and other risks, so too should students be protected against poor handling of their sensitive information. This does not just include research done in an academic setting, which is covered by our research exemption clause and by existing law, but also informal research done by data scientists within a MOOC provider. Such a clause on access control could help ensure that this informal research is held to similar standards.

Following Principle 4, Educational Use and Context, operators should publish their data pipelines for students to see and to ensure them that their data is being used by responsible parties for educational purposes. This should include the types of data shared internally and externally, the number of personnel at each access level, and the training or other appropriate qualification required and received by said personnel.

> *(9) An operator must publish its data pipeline to students, including but not limited to*
>
> *(a) what types of data are shared with personnel;*
> *(b) the number of personnel at each access level; and*
> *(c) training or qualification required and received at each access level, including current employment status or other criteria for terminating access.*
>
> *Furthermore, an operator must implement and publish security measures to ensure that*
>
> *(d) only authorized individuals may access particular types of data, where authorization is established as in (9)(a-c); and*
> *(e) an individual is immediately denied access to data when authorization is withdrawn.*

Enforcement and Redress

The final and potentially most important part of our proposed legislation is that of enforcement and redress. The FIPPs note that it is "generally agreed that the core principles of privacy protection can only be effective if there is a mechanism in place to enforce them."[157] They outline three types of enforcement and redress: self-regulation, private remedies, and government enforcement.[158]

While some edtech companies currently self-regulate through agreements like the Student Privacy Pledge, these agreements are voluntary to enter into and don't include the Big Three in MOOCs.[159] In order to protect students according to Principles 1 through 4,

[156] Office for Human Research Protections, "45 CFR 46."
[157] Federal Trade Commission, "Priavcy Online: A Report to Congress," 10.
[158] Ibid., 10-11.
[159] Student Privacy Pledge, "Signatories."

stronger enforcement is needed.

Operators should be held liable through a private right of action. This is line with the FIPPs' private remedies and EPIC's suggestion for additions to SOPIPA. It also offers redress to the affected students.[160] [161] Additionally, in cases where students may not request an action for injunctive relief or damages, the Attorney General and the FTC should be able to impose penalties and issue warnings on behalf of the government. These enforcement provisions draw upon provisions from the Illinois School Student Records Act and the Idaho Student Data Accessibility, Transparency and Accountability Act of 2014.[162] [163]

> *(10) Any operator determined to have violated a provision from sections (1) through (9) shall be liable for a civil penalty not to exceed one hundred thousand dollars ($100,000) per violation. The penalty shall be imposed by a civil enforcement action brought by the Attorney General or the Federal Trade Commission in the district court for the district in which the violation occurred. The penalty may be reduced at the discretion of the court if the operator can demonstrate that (a) the violation caused no substantial harm to students and (b) efforts to comply were promptly taken upon notice of the violation. Any students affected by the violation shall be notified in a timely manner.*
>
> *(11) Any student injured or adversely affected by an operator's violation of a provision from sections (1) through (9) may request both an action for injunctive relief and an action for damages in the district court for the district in which the student is located or in which the operator is located. In the case of a successful action, the operator is liable to the plaintiff for the plaintiff's damages and cost of action, including attorney fees, as determined by the court.*

5.5 Reactions and Implications

We now consider the effects that a Student Internet Privacy Act would have on various stakeholders, as well as how they might react to such a proposal.

Students

Students will likely be strong supporters of our proposed legislation because of the privacy protections it grants them. In particular, the proposed regulations allow MOOCs to continue providing value to students while balancing the need for data privacy.

We chose to include all MOOC users as students because the users enroll with similar goals as students in traditional schools, such as accessing high-quality education that could

[160] Federal Trade Commission, "Priavcy Online: A Report to Congress," 10-11.
[161] EPIC, "State Student Privacy Policy."
[162] Illinois General Assembly, "(105 ILCS 10/) Illinois School Student Records Act."
[163] California Senate, "Senate Bill No. 1177 Chapter 839," 22584(e)(2).

increase their chances to be hired by employers. Our policy recommendations, if enforced, provide students on MOOCs with privacy protections similar to those granted by FERPA. Parents and students have made it clear that they want transparency, consent, and control with regards to collected student data and how it is used.[164] These basic privacy expectations, acknowledged within OECD's privacy guidelines[165] and the FTC's FIPPs,[166] will allow online students to feel secure from harmful data practices that could jeopardize their future opportunities.

Our proposed bill's clauses that prohibit targeted advertising, student profiles, and the sale of student data also benefit students by ensuring that no third party receives their data without their knowledge. They further protect students against biases that could ensue when seeking jobs or applying to other institutions like college.[167] The ability to view a data dashboard, as well as delete, amend, or export one's data, provides students with control over their data. All of these provisions are likely to be favored by students for both privacy considerations and other reasons.

However, preventing targeted advertising could limit the benefits that students would gain from personalized recommendations. For example, targeted advertisements that include recommended books or other websites that may help a student with a particular subject or topic would be banned by our recommendations. We believe this is a small downside and of low priority compared to the possibility of opening up opportunities for misuse of student data that can ensue from allowing targeted advertising. Additionally, prohibiting targeted advertising would serve students by preventing distractions while they are learning. As the Washington Post writes, "Students are also subjected to numerous ads via YouTube and the other websites assigned by the platform, which can be very distracting, especially for children with special needs."[168]

Academic Researchers

Researchers will not be affected significantly by our proposed legislation, so their response is likely to be neutral. They will be strong supporters of the research exemption clause, which allows academic research to continue largely unimpeded and without requirements of de-identification.

We chose to include the research exception clause because studies involving human subjects are well-regulated, and they provide valuable insights without causing major privacy harms. One of the primary benefits of MOOCs is the creation of large datasets that can provide rich research material on many levels, as detailed in Section 2.

[164] Idaho Legislature, "Senate Bill No. 1296."
[165] OECD, "The OECD Privacy Framework."
[166] Federal Trade Commission, "Privacy Online: A Report to Congress."
[167] Jones, Meg Leta and Lucas Regner, "Users or Students? Privacy in University MOOCs," 1493.
[168] Strauss, Valerie, "Parents cite student privacy concerns with popular online education platform."

However, the proposed legislation does create a few additional implementation constraints for researchers to consider when planning studies. First, mandating that MOOCs allow students to delete their accounts and associated data may concern researchers. For the most part, these concerns are unfounded as the effects on research are contained. If a student decides to delete her account midway through a course, then the data that is missing from the second half would not have existed even without the data deletion provision. Additionally, our proposal lets researchers keep the students' data for research purposes even if the MOOC platform is required to delete its copies, so long as the collection mechanisms were specified and agreed to in advance. The one way that researchers might be affected is that they would now be incentivized to collect information on a regular and ongoing basis from MOOCs—a potentially perverse incentive. This is because if they wait too long to request data, students may have already deleted their accounts and associated data, and such omissions would lead to self-selection bias in research.

Additionally, if a study collects special information or otherwise requires unique data processing, researchers may need to work with the MOOC providers to modify the data dashboard and data export function to ensure students have the same access as the researchers. However, given the ease and relatively low overhead with which software is updated for minor process changes—as well as the consideration that edX largely complies with the proposed legislation already—we do not expect any substantial chilling effects on research. We believe researchers will appreciate the protections that the legislation provides to students, and that they will find these tradeoffs to be acceptable.

MOOC Providers

MOOC providers are likely to oppose our proposed legislation because it limits certain student data monetization strategies and requires multiple new feature implementations. However, some MOOCs might support the proposal, particularly those which are non-profit and prioritize improving access to education over profit generation.

Considering the various business models discussed earlier, our proposed legislation will likely have a chilling effect on each model, excluding the traditional "school-like" certification model. Similar to SOPIPA, we prevent MOOC providers from monetizing student data through selling it to marketers and advertisers, stifling targeted advertising as a revenue stream. In the job matching model, MOOCs may still offer job matching services to students and employers, but they will no longer be able to offer student profiles wholesale to paying companies. Furthermore, our bill's introduction of access control will require MOOCs to expend resources in software and business infrastructure changes. Such restrictions might force some MOOCs, particularly those small in size and those that solely depend on commercial use of data, to shut down.

However, we believe the proposed legislation is desperately needed and may in fact help MOOCs in a couple long-term ways. First, our proposal forces MOOCs to practice

transparency, which fosters trust and healthy relationships between providers and users. Without it, companies may end up entangled in public relation nightmares like in the case of inBloom. Second, when MOOCs are held accountable for protecting student data, they are also incentivized to produce high-quality content. This is because their content must be designed to increase student participation as opposed to cutting corners and simply exploiting student data for commercial uses. Finally, note that our proposal would not ban MOOCs from collecting data in general. MOOCs could still use data to improve their teaching methods and to market key performance indicators in aggregate.

Universities

Universities will likely react neutrally as they are not seriously harmed nor helped by our proposed legislation. Within universities, academics are likely to be supportive because they are naturally inclined to be in favor of student protections.

However, one consideration for universities is the additional burden of implementing features on MOOCs run by universities or with which universities integrate. Our proposal requires MOOCs to support amendment, export, and deletion features. Because amending and deleting data are inherently part of any functional MOOC, extending the permission structure and university processes to afford these capabilities to students should not be terribly difficult. Ultimately, universities will likely be pleased with the clarity the proposed legislation brings to their online activities, and perhaps the savings in legal fees will cover the costs of software implementation.

Just as with business models for MOOC providers, future directions for university growth are somewhat restricted by our proposal. For example, MIT runs one of the largest annual student-run career fairs in the country. Employers pay large sums for booth space and access to a student resume database. This results in over a million dollars of revenue to the groups that run the fair. Career fair models may be a useful avenue to augment universities' coffers, but the proposed legislation disallows certain types of resume or skill databases if they pull data from MOOC platforms. Similarly, any other business model a university might wish to pursue online would likely be subject to the same restrictions as the MOOC providers themselves. While universities might have concerns over such limitations, they are likely to be dampened by the fact that most universities already have well established business models to support their activities.

U.S. Attorney General and FTC

The U.S. Attorney General and Federal Trade Commission will likely support our proposed legislation, given that similar laws at the state level are already being successfully enforced by their state equivalents. The main issue is whether or not the proposal provides adequate enforcement power, and indeed, it does.

Similar to SOPIPA being enforced by California's Attorney General as part of their Unfair Competition Law, our bill would charge the United States Attorney General and the FTC with enforcing any enacted rules. The U.S. Attorney General and FTC—which have a history of working together[169]—would treat any violations by MOOC providers as unlawful business practices, and it would be in their power to prosecute the provider and to impose a penalty. There is even precedent for such scenarios at the state level. For instance, earlier this year, Mississippi's Attorney General filed a lawsuit against Google over student privacy.[170] Given that several state attorneys general support state laws like SOPIPA currently protecting student privacy, we expect that the U.S. Attorney General and FTC would not face problems while similarly enforcing our federal policy recommendations.

However, the Attorney General and FTC do have the potential to become overburdened by enforcing violations. Furthermore, the individuals harmed may not obtain compensation. For these reasons, we included the private right of action as another outlet that allows for enforcement without the Attorney General or FTC.

Department of Education

The Department of Education (ED) will likely have a neutral or positive reaction to our proposed legislation because it protects students' rights. However, it would not be enforcing the new rules under the proposed legislation.

Our reason for not choosing ED to enforce the rules is that ED's mandate is to focus particularly on schools when "prohibiting discrimination" and "distributing as well as monitoring" financial aid.[171] Because K–12 edtech companies like Khan Academy and MOOCs like Udacity fall outside the scope of ED, it would be unreasonable to burden ED with enforcement of our Student Internet Privacy Act.

In the future, it is possible that the structure and focus of ED may change so that the digital intersection of schools, edtech companies, and MOOC providers fall within the purview of ED. However, at the time of writing, it seems prudent to call upon the United States Attorney General and tort system to enforce the rules, allowing ED to focus on its current stated mission.

Congress

It is difficult to predict the reaction of Congress, but we believe the policy environment is becoming more receptive and therefore favorable towards our proposed legislation. The bill would likely be bipartisan, although the largest obstacle would be conservative members' concerns over the potential chilling effects on businesses.

[169] Federal Trade Commission, "The Enforcers."
[170] Herold, Benjamin, "Missisippi Attorney General Sues Google Over Student-Data Privacy."
[171] ED, "Overview and Mission Statement."

There are a number of members of Congress who have already shown support for student privacy online. Senators Edward J. Markey (D-MA) and Orrin Hatch (R-UT) have co-sponsored the Protecting Student Privacy Act multiple times in the hopes of updating FERPA for the digital age.[172] The Student Digital Privacy and Parental Rights Act of 2015 was itself a bipartisan effort by Representatives Jared Polis (D-CO) and Luke Messer (R-IN).[173] A similar bill was proposed in the Senate by Senators Richard Blumenthal (D-CT) and Steve Daines (R-MT).[174] All six of these members of Congress are likely to support a bill that protects the privacy rights of all online students, including MOOC users.

The congressmen would be bolstered by increasing public support. According to Pew Research Center, some studies show "Americans are becoming more anxious about their privacy, especially in the context of digital technologies that capture a wide array of data about them."[175] This past October, for instance, Mattel reversed its plan to sell a smart device for children called Aristotle after "child advocacy groups, lawmakers and parents raised concerns"[176] about its impact on children's privacy and development.[177] States have already recognized this and are passing many new privacy bills.

However, it is important to note that the United States Congress has failed to pass student digital privacy legislation in prior years. Many conservatives may prefer to allow states to continue enacting their own laws, as opposed to an overarching federal law. This in fact was one of the criticisms that led to Polis and Messer's bill not progressing through Congress in 2015.[178] Further criticism would likely come from strong libertarians such as Senator Rand Paul (R-KY), who might fear chilling effects on MOOC and edtech business models. It is promising to note, however, that only 11% of adults believe privacy regulations "would be overly burdensome and stifle innovation" and "hurt the people they're intended to help by making educational tools more expensive."[179]

6. Conclusion

We summarize the main thrusts of the paper by considering the key questions we have tried to answer.

Why now?

Distance learning has been around for centuries, as has privacy-protecting legislation. However, new technologies like the internet, new data-processing capabilities like machine

[172] Markey.senate.gov, "Senators Markey & Hatch Reintroduce Bipartisan Legislation to Protect Student Privacy."
[173] Singer, Natasha, "Legislators Introduce Student Digital Privacy Bill."
[174] Ujifusa, Andrew, "Education Data, Student Privacy Take Spotlight at Capitol Hill Hearing."
[175] Rainie, Lee and Shiva Maniam, "Americans feel the tensions between privacy and security concerns."
[176] Peachman, Rachel Rabkin, "Mattel Pulls Aristotle Children's Device After Privacy Concerns."
[177] Ibid.
[178] Herold, Benjamin, "Federal Student-Data-Privacy Bill Delayed Following Criticism."
[179] Common Sense Media, "Student Privacy Survey."

learning, and new business models like MIT's MicroMasters have created a paradigm shift in education from the physical, standardized classroom to the digital, individually-tailored MOOC. It is clear that existing legislation does not adequately protect student privacy in this new context. There is a unique window of opportunity to introduce legislation now before students get hurt as companies in the future seek to out-compete each other through dubious privacy-infringing practices which are admissible under current law.

What benefits do MOOCs provide?

MOOCs provide widespread access to higher education as well as new research opportunities. Principal factors driving access include internet-based remote content, free or low-cost instruction and certification from elite institutions, and flexible course formats and schedules. Factors driving research include large-scale datasets, fine-grain data collection, diversity of MOOC participants, and access to personal information subject to applicable research regulations.

What privacy issues do MOOCs raise?

MOOCs take on many roles traditionally filled by schools, so it is natural that they have intimate contact with student data. However, the digital nature of the MOOC enables unprecedented collection of vast quantities of data on individual students. This large-scale collection, coupled with the expansion of what constitutes personally identifiable information, creates difficulties governing the appropriate collection and use of student data. Data scientists are able to extrapolate much more about a student's personal life and behavior than ever before. Business models that are currently acceptable in other contexts can lead to student privacy infringement when carried over to the context of education. Commercial entities that supplant schools are not inherently bad, but they are prone to follow weaker privacy protection protocols. Most concerning is the potential for MOOC providers to exploit or otherwise capitalize upon their data assets to the detriment of student privacy with targeted advertising. These issues can be fixed with increased legislative focus on transparency and data control, but student disempowerment remains admissible under current law.

How effective is existing legislation?

Existing legislation provides helpful guidance but ultimately falls short of protecting students' privacy. At the national level, FERPA is appropriate in spirit but contains ineffective enforcement measures, a poor definition of "school officials," and only applies to schools receiving federal funds. At the state level, California's SOPIPA addresses some gaps in FERPA; notably, it applies to any K–12 edtech company. However, it fails to protect learners outside K–12, to give students control over their information, to provide transparency around data collected, and to establish strong enforcement guidelines. The Student Privacy Pledge is a voluntary K–12 edtech pact which focuses more on

transparency but still fails to provide strong data control measures and especially enforcement. Other state laws have since been passed in the spirit of SOPIPA, but they too fail to address the main pitfalls in the California legislation.

What recommendations might help address privacy concerns?

We propose the Student Internet Privacy Act, legislation at the national level that would apply to all digital services—including MOOCs—that serve the same primary functions as traditional schools. The standard definition of PII is taken directly from SOPIPA, and the four principles of transparency, data control, focused collection, and educational use and context guide our recommendations for key provisions in our proposed law: no targeted advertising, no student profiles except for educational uses, no sale of student information, research exemption, data dashboards, account deletion, data amendment, data export, access control, and enforcement and redress. The proposed legislation directly addresses the concerns and gaps from existing legislation, while expanding the scope to cover all students.

How do the recommendations affect various stakeholders?

Positive or neutral reactions are expected of most stakeholders, including students, researchers, universities, and relevant government agencies. The largest concern from MOOC providers and libertarian supporters of the free market is with chilling effects on MOOC business models: the recommendations stifle potential revenue streams. However, it must be noted that most MOOCs successfully exist at present without revenue from exploitative business models, and implementing the recommendations is not expected to be especially burdensome. The group most supportive of the Student Internet Privacy Act will not surprisingly be students, as their online privacy rights would finally be protected in a comprehensive way under this legislation.

Acknowledgements and Contributions

The authors would like to thank several people without which this report would not have been possible. Michael Trice went above and beyond to guide us in developing a unified paper with a strong recommendation, meeting with us outside of work hours. Daniel Weitzner urged us to think beyond FERPA and to first develop a set of principles to guide any such legislation. Jon Daries graciously volunteered an hour of his time to give us a crash course on the nitty-gritty details of working with student data from edX within a university.

Among the authors, our work was distributed evenly. Drew Bent focused on researching the current regulatory landscape for MOOCs, identifying shortcomings and drafting the proposed legislation. Nitah Onsongo worked on understanding MOOC providers' business models and the commercial uses of student data. Andy Trattner investigated how research is conducted both in academic environments and inside MOOC providers. We worked together in person and remotely to write the paper, and for finishing touches, Andy nitpicked wordings as well as compiled our footnotes and Works Cited, Drew chose a lovely font and edited to ensure a consistent voice, and Nitah performed rigorous copy editing across the paper.

Works Cited

Alim, Frida, Nate Cardozo, Gennie Gebhart, Karen Gullo, and Amul Kalia. 2017. "Spying on Students." *Electronic Frontier Foundation*. Accessed 28 November 2017. https://www.eff.org/files/2017/04/13/student-privacy-report.pdf.

Bayeck, Rebecca Yvonne. 2016. "Exploratory study of MOOC learners' demographics and motivation: The case of students involved in groups." *Open Praxis*, vol. 8, no. 3, pp 223-233. https://oerknowledgecloud.org/sites/oerknowledgecloud.org/files/282-1432-2-PB.pdf.

Belleflamme, Paul and Julien Jacqmin. 2016. "An Economic Appraisal of MOOC Platforms: Business Models and Impacts on Higher Education." *CESifo Economic Studies*, vol. 62, no. 1, pp 148-169.

Bowden, Tracy. 2014. "Google Apps for Education mining data to develop targeted ads, experts warn." *ABC News*. Accessed 28 November 2017. http://www.abc.net.au/news/2014-06-18/cyber-experts-say-google-data-mining-schools-for-targeted-ads/5533752.

California Senate. 2014. "Senate Bill No. 1177 Chapter 839." Accessed 27 November 2017. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201320140SB1177.

Chen, Zhongzhou, Christopher Chudzicki, Daniel Palumbo, Giora Alexandron, Youn-Jeng Choi, Qian Zhou, and David E. Pritchard. 2016. "Researching for better instructional methods using AB experiments in MOOCs: results and challenges." *Research and Practice in Technology Enhanced Learning*, vol 11, no. 9. Springer.

Chuang, Isaac and Andrew Ho. 2016. "HarvardX and MITx: Four Years of Open Online Courses -- Fall 2012-Summer 2016." *Social Science Research Network*. Last revised 16 January 2017. Accessed 27 November 2017.

Common Sense Media. 2014. "Student Privacy Survey." Accessed 28 November 2017. https://www.commonsensemedia.org/sites/default/files/uploads/about_us/student_privacy_survey.pdf.

Congress.gov. 2015. "H.R. 2092." Accessed 28 November 2017. https://www.congress.gov/114/bills/hr2092/BILLS-114hr2092ih.pdf.

Congress.gov. 2015. "H.R.2092 - Student Digital Privacy and Parental Rights Act of 2015." *Library of Congress*. Accessed 28 November 2017. https://www.congress.gov/bill/114th-congress/house-bill/2092.

Crunchbase. 2017. "Coursera." Accessed 27 November 2017. https://www.crunchbase.com/organization/coursera.

Daries, Jon P., Justin Reich, Jim Waldo, Elise M. Young, Jonathan Whittinghill, Andrew Dean Ho, Daniel Thomas Seaton, and Isaac Chuang. 2014. "Privacy, Anonymity, and Big Data in the Social Sciences." *Communications of the ACM*, vol. 57, no. 9, pp 56-63.

Data Quality Campaign. 2016. "Student Data Privacy Legislation: A Summary of 2016 State Legislation." Accessed 28 November 2017. https://dataqualitycampaign.org/wp-content/uploads/2016/09/DQC-Legislative-summary-09302016.pdf.

Decarr, Kristin. 2017. "Google Changes Mind, Signs Obama's Student Privacy Pledge." *Education News*. Accessed 28 November 2017. http://www.educationnews.org/technology/google-changes-mind-signs-obamas-privacy-pledge/.

DeNisco, Alison. "Protecting student data in the age of marketing and advertising." *CIO News*. Accessed 5 December 2017. https://www.districtadministration.com/article/protecting-student-data-age-marketing-and-advertising

Dewey, Caitlin. 2016. "98 personal data points that Facebook uses to target ads to you." *The Washington Post*. Accessed 27 November 2017. https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/.

Diaz-Hernandez, Ana. 2015. "Announcing Talent Source: Connecting Students & Employers With Our New Candidate Sourcing Tool." *Udacity*. Accessed 27 November 2017. https://blog.udacity.com/2015/10/announcing-talent-source-connecting-students-employers-with-our-new-candidate-sourcing-tool.html.

Doran, Leo. 2016. "Ambitious Student-Data-Privacy Law in California Attracts National Attention." *Education Week*. Accessed 28 November 2017. http://blogs.edweek.org/edweek/DigitalEducation/2016/01/ambitious_student-data-privacy.html.

edX. 2017. "EdX Research Guide." Accessed 27 November 2017. http://edx.readthedocs.io/projects/devdata/en/latest/index.html.

edX. 2014. "Privacy Policy." Last updated 22 October 2014. https://www.edx.org/edx-privacy-policy.

Electronic Privacy Information Center. n.d. "Family Educational Rights and Privacy Act (FERPA)." Accessed 27 November 2017. https://epic.org/privacy/student/ferpa/.

Electronic Privacy Information Center. n.d. "State Student Privacy Policy." Accessed 27 November 2017. https://epic.org/state-policy/student-privacy/.

Electronic Privacy Information Center. n.d. "Student Privacy Bill of Rights." Accessed 28 November 2017. https://epic.org/privacy/student/bill-of-rights.html.

Federal Trade Commission. n.d. "The Enforcers." Accessed 8 December 2017. https://www.ftc.gov/tips-advice/competition-guidance/guide-antitrust-laws/enforcers.

Federal Trade Commission. June 1998. "Priavcy Online: A Report to Congress." Accessed 28 November 2017. https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf.

Fischer, Jill. 2017. "Top 4 FAQ about the MicroMasters Credential." *edX*. Accessed 27 November 2017. https://blog.edx.org/top-4-faq-micromasters-credential.

Fitzpatrick, Kaleigh C. 2016. "Student Data at Risk: A Multi-Tiered Approach for Massachusetts to Mitigate Privacy Risks While Utilizing Innovative Education Technology in Schools." *Journal of High Technology Law*, vol. 16, no. 1.5, pp 294-339.

Foxman, Maxwell, Mateescu Alexandra, and Bulger Monica. "Advertising in Schools." Accessed 05 December 2017. https://datasociety.net/pubs/ecl/Advertising_primer_2016.pdf

Friedlander, Simone A. "Net Neutrality and the FCC's 2015 Open Internet Order." *Berkeley Technology Law Journal*. Annual Review 2016 (2016): 905. *HeinOnline*, EBSCOhost.

Future of Privacy Forum. 2016. "FPF Guide to Protecting Student Data Under SOPIPA." Accessed 28 November 2017. https://fpf.org/wp-content/uploads/2016/11/SOPIPA-Guide_Nov-4-2016.pdf.

Georgia Tech. n.d. "Why OMS CS?" Accessed 27 November 2017. https://www.omscs.gatech.edu/explore-oms-cs.

Google. 2013. "Google Books Ngram Viewer." Accessed 27 November 2017. https://books.google.com/ngrams.

Haimson, Leonie and Rachael Stickland. n.d. "About Us." *Parent Coalition for Student Privacy*. Accessed 27 November 2017. https://www.studentprivacymatters.org/about-us/.

Halpert, Jim and Michelle Anderson. 2015. "State Privacy and Security Developments–Looking Back and Looking Ahead." *Bloomberg Law*. Accessed 27 November 2017. https://www.bna.com/state-privacy-security-n17179922907/.

Herold, Benjamin. 2015. "Draft of President Obama's Student-Data-Privacy Bill Raises Questions." *Education Week*. Accessed 28 November 2017. http://blogs.edweek.org/edweek/DigitalEducation/2015/01/federal_student-data-privacy_draft_bill.html.

Herold, Benjamin. 2015. "Federal Student-Data-Privacy Bill Delayed Following Criticism." *Education Week*. Accessed 28 November 2017. http://blogs.edweek.org/edweek/DigitalEducation/2015/03/federal_student_data_privacy_bill_delayed.html.

Herold, Benjamin. 2017. "Missisippi Attorney General Sues Google Over Student-Data Privacy." *Education Week*. Accessed 28 November 2017. http://blogs.edweek.org/edweek/DigitalEducation/2017/01/mississippi_sues_google_student_data_privacy.html.

Harris, Kamala D. 2016. "Ready for School." *Office of the Attorney General, California Department of Justice*. Accessed 27 November 2017. https://oag.ca.gov/sites/all/files/agweb/pdfs/cybersecurity/ready-for-school-1116.pdf.

Holmberg, Börje. "The Evolution, Principles and Practices of Distance Education." In *Studien und Berichte der Arbeitsstelle Fernstudienforschung der Carl von Ossietzky Universität Oldenburg*, vol. 11, p 13. *Bibliotheks-und Informationssystem der Universitat Oldenburg*, 2005.

Hudson, David. 2015. "The President Announces New Actions to Protect Americans' Privacy and Identity." *Obama White House Archives*. https://obamawhitehouse.archives.gov/blog/2015/01/12/president-announces-new-actions-protect-americans-privacy-and-identity.

Idaho Legislature. 2014. "Senate Bill No. 1296." Accessed 28 November 2017. https://legislature.idaho.gov/wp-content/uploads/sessioninfo/2014/legislation/S1296.pdf.

Illinois General Assembly. n.d. "(105 ILCS 10/) Illinois School Student Records Act." Accessed 28 November 2017. http://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=1006&ChapterID=17.

Jones, Meg Leta and Lucas Regner. 2016. "Users or Students? Privacy in University MOOCs." *Science and Engineering Ethics*, vol. 22, no. 5, pp 1473-1496. https://link.springer.com/article/10.1007/s11948-015-9692-7.

Fischer, Gerhard. 2014. "Beyond hype and underestimation: identifying research challenges for the future of MOOCs." *Distance Education*, vol. 35, no. 2, pp 149-158.

LII. 2017. "34 CFR 99.31 - Under what conditions is prior consent not required to disclose information?" Accessed 27 November 2017. https://www.law.cornell.edu/cfr/text/34/99.31.

Markey.senate.gov. 2017. "Senators Markey & Hatch Reintroduce Bipartisan Legislation to Protect Student Privacy." Accessed 28 November 2017. https://www.markey.senate.gov/news/press-releases/senators-markey-and-hatch-reintroduce-bipartisan-legislation-to-protect-student-privacy.

Messer, Luke. 2015. "Student Digital Privacy and Parental Rights Act of 2015." Accessed 28 November 2017. https://polis.house.gov/uploadedfiles/messer_polis_student_data_privacy_final.pdf.

MIT Institutional Research. 2014. "MITx Data Request Checklist." Last updated August 2014. http://web.mit.edu/ir/mitx/.

MyLeanMBA. 2017. "Breaking down the top 3 MOOC platforms: Coursera, Udacity & edX." *Medium*. Accessed 27 November 2017. https://medium.com/@MyLeanMBA/breaking-down-the-top-3-mooc-platforms-coursera-udacity-edx-13e5ed481337.

Nasseh, Bizhan. 1997. "A Brief History of Distance Education." *SeniorNet*. Accessed 27 November 2017. http://www.seniornet.org/edu/art/history.html.

OECD. 2013. "The OECD Privacy Framework." Accessed 27 November 2017. http://www.oecd.org/sti/ieconomy/privacy.htm.

Office for Human Research Protections. "45 CFR 46." *U.S. Department of Health and Human Services*. Last revised 15 January 2010. https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html.

Open University, The. "The OU Story." Accessed 27 November 2017. http://www.open.ac.uk/about/main/strategy/ou-story.

Parent Coalition for Student Privacy. 2015. "Five Principles to Protect Student Data Privacy." Accessed 7 December 2017. https://www.studentprivacymatters.org/five-principles-to-protect-student-data-privacy/.

Parent Coalition for Student Privacy. n.d. "State Student Privacy Laws." Accessed 28 November 2017. https://www.studentprivacymatters.org/state-legislation/.

Peachman, Rachel Rabkin. 2017. "Mattel Pulls Aristotle Children's Device After Privacy Concerns." *The New York Times*. Accessed 28 November 2017. https://www.nytimes.com/2017/10/05/well/family/mattel-aristotle-privacy.html.

Penrose, Mary Margaret. 2011. "In The Name Of Watergate: Returning FERPA To Its Original Design." *N.Y.U. Journal of Legislation & Public Policy*, vol. 14, no. 1, pp 75-113.

Peterson, Dylan. 2016. "Edtech and Student Privacy: California Law as a Model." *Berkeley Technology Law Journal*, Special Issue, vol. 31, pp 961-995.

Polonetsky, Jules and Omer Tene. 2015. "Who is Reading Whom Now: Privacy in Education from Books to MOOCs." *Vanderbilt Journal of Entertainment and Technology Law*, vol. 17, no. 4, pp 927-990.

Privacy Technical Assitance Center (PTAC). 2014. "Protecting Student Privacy While Using Online Educational Services: Requirements and Best Practices." Accessed 27 November 2017. https://tech.ed.gov/wp-content/uploads/2014/09/Student-Privacy-and-Online-Educational-Services-February-2014.pdf.

Rainie, Lee and Shiva Maniam. 2016. "Americans feel the tensions between privacy and security concerns." *Pew Research Center*. Accessed 28 November 2017. http://www.pewresearch.org/fact-tank/2016/02/19/americans-feel-the-tensions-between-privacy-and-security-concerns/.

Roscorla, Tanya. 2016. "3 Student Data Privacy Bills That Congress Could Act On." *The Center for Digital Education*. Accessed 28 November 2018. http://www.centerdigitaled.com/K–12/3-Student-Data-Privacy-Bills-That-Congress-Could-Act-On.html.

Seaton, Daniel, Cody Coleman, Jon Daries, and Isaac Chuang. 2015. "Enrollment in MITx MOOCs: Are We Educating Educators?" *Educause*. Accessed 27 November 2017. https://er.educause.edu/articles/2015/2/enrollment-in-mitx-moocs-are-we-educating-educators.

Shah, Dhawal. 2017. "Massive List of MOOC Provides Around The World." *Class Central*. Accessed 27 November 2017. https://www.class-central.com/report/mooc-providers-list/.

Singer, Natasha. 2015. "Legislators Introduce Student Digital Privacy Bill." *New York Times*. Accessed 28 November 2017. https://bits.blogs.nytimes.com/2015/04/29/legislators-introduce-student-digital-privacy-bill/.

Solove, Daniel. 2012. "FERPA and the Cloud: What FERPA Can Learn from HIPAA." *LinkedIn*. Accessed 27 November 2017. https://www.linkedin.com/pulse/20121218131535-2259773-ferpa-and-the-cloud-what-ferpa-can-learn-from-hipaa/.

Solove, Daniel J. and Woodrow Hartzog. 2014. "The FTC and the New Common Law of Privacy." *Columbia Law Review*, vol. 114, pp 583-676.

Solove, Daniel J. 2006. "A Taxonomy of Privacy." *University of Pennsylvania Law Review*, vol. 154, no. 3, pp 477-560. Last revised 6 May 2008.

Straumsheim, Carl. 2017. "MIT Deems MicroMasters a Success." *Inside Higher Ed*. Accessed 27 November, 2017. https://www.insidehighered.com/news/2017/07/26/mit-deems-half-online-half-person-masters-program-success.

Straumsheim, Carl. 2015. "MOOCs for (a Year's) Credit." *Inside Higher Ed*. Accessed 27 November, 2017. https://www.insidehighered.com/news/2015/04/23/arizona-state-edx-team-offer-freshman-year-online-through-moocs.

Strauss, Valerie. 2013. "Lawsuit charges Ed Department with violating student privacy rights." *The Washington Post*. Accessed 28 November 2017. https://www.washingtonpost.com/news/answer-sheet/wp/2013/03/13/lawsuit-charges-ed-department-with-violating-student-privacy-rights/.

Strauss, Valerie. 2015. "The astonishing amount of data being collected about your children." *The Washington Post*. Accessed 27 November 2017. https://www.washingtonpost.com/news/answer-sheet/wp/2015/11/12/the-astonishing-amount-of-data-being-collected-about-your-children/.

Strauss, Valerie. 2017. "Parents cite student privacy concerns with popular online education platform." *The Washington Post*. Accessed 28 November 2017. https://www.washingtonpost.com/news/answer-sheet/wp/2017/08/30/parents-cite-student-privacy-concerns-with-popular-online-education-platform/.

Student Privacy Pledge. n.d. "FAQs." *Future of Privacy Forum* and *The Software & Information Industry Association*. Accessed 28 November 2017. https://studentprivacypledge.org/faqs/.

Student Privacy Pledge. n.d. "Privacy Pledge." *Future of Privacy Forum* and *The Software & Information Industry Association*. Accessed 28 November 2017. https://studentprivacypledge.org/privacy-pledge/.

Student Privacy Pledge. n.d. "Signatories." *Future of Privacy Forum* and *The Software & Information Industry Association*. Accessed 28 November 2017. https://studentprivacypledge.org/signatories/.

Student Privacy Pledge. 2016. "Student Privacy Pledge Reaches Milestone of 300 Signatories." *Future of Privacy Forum* and *The Software & Information Industry Association*. Accessed 28 November 2017. https://studentprivacypledge.org/student-privacy-pledge-reaches-milestone-of-300-signatories/.

Thomas, Gail S. 1988. "Connected Education, Inc." *Netweaver*. Electronic Networking Association. Accessed 27 November 2017. https://web.archive.org/web/20080827214245/http://cgi.gjhost.com/~cgi/mt/netweaverarchive/000144.html.

Ujifusa, Andrew. 2016. "Education Data, Student Privacy Take Spotlight at Capitol Hill Hearing." *Education Week*. Accessed 28 November 2017. http://blogs.edweek.org/edweek/campaign-K–12/2016/03/education_privacy_research_congressional_hearing.html.

U.S. Department of Education. 2017. "About Us." Accessed 27 November 2017. https://studentprivacy.ed.gov/about-us.

U.S. Department of Education. n.d. "Overview and Mission Statement." Accessed 28 November 2017. https://www2.ed.gov/about/landing.jhtml.

Virginia General Assembly. 2016. "HB 749." *Virginia's Legislative Information System*. Accessed 8 December 2017. https://lis.virginia.gov/cgi-bin/legp604.exe?161+sum+HB749.

White House Office of the Press Secretary, The. 2015. "Remarks by the President at the Federal Trade Commission." Accessed 28 November 2017.

https://obamawhitehouse.archives.gov/the-press-office/2015/01/12/remarks-president-federal-trade-commission.

White House Office of the Press Secretary, The. 2012. "We Can't Wait: Obama Administration Unveils Blueprint for a 'Privacy Bill of Rights' to Protect Consumers Online." Accessed 28 November 2017. https://obamawhitehouse.archives.gov/the-press-office/2012/02/23/we-can-t-wait-obama-administration-unveils-blueprint-privacy-bill-rights.

Appendix: Draft Legislation

A BILL

To require operators that provide online and similar services to educational institutions, namely K–12 educational software and MOOCs, to protect the privacy of students.

SECTION 1. SHORT TITLE.

This Act may be cited as the "Student Internet Privacy Act".

SEC. 2. DEFINITIONS.

(a) **IN GENERAL.**—In this Act:

(1) **OPERATOR.**—The term "operator" means either a K–12 edtech operator or MOOC operator, given the following definitions

(A) "K–12 edtech operator": the operator of an Internet Web site, online service, online application, or mobile application with actual knowledge that the site, service, or application is used primarily for K–12 school purposes and was designed and marketed for K–12 school purposes

(B) "MOOC operator": the operator of an Internet Web site, online service, online application, or mobile application that serves the same primary functions as traditional schools, including courseware and some form of certification

(2) **COVERED INFORMATION.**—The term "covered information" means personally identifiable information or materials, in any media or format that meets any of the following:

(A) Is created or provided by a student, or the student's parent or legal guardian, to an operator in the course of the student's, parent's, or legal guardian's use of the operator's site, service, or application for educational purposes.

(B) Is created or provided by an employee or agent of the K–12 school, school district, local education agency, or county office of education, to an operator.

(C) Is gathered by an operator through the operation of a site, service, or application described in (1) and is descriptive of a student or otherwise identifies a student, including, but not limited to, information in the student's educational record or email, first and last name, home address, telephone number, email address, or other information that allows physical or online contact, discipline records, test results, special education data, juvenile dependency records, grades, evaluations, criminal records, medical records, health records, social security number,

biometric information, disabilities, socioeconomic information, food purchases, political affiliations, religious information, text messages, documents, student identifiers, search activity, photos, voice recordings, or geolocation information.

(3) **EDUCATIONAL PURPOSES.**—The term "educational purposes" means purposes that customarily take place at the direction of the K–12 or university school, teacher, or school district or aid in the administration of school activities, including, but not limited to, instruction in the classroom or at home, administrative activities, and collaboration between students, school personnel, or parents, or are for the use and benefit of the school

(4) **STUDENT.**—The term "student" means any user of the operator's Internet Web site, online service, online application, or mobile application.

(5) **DATA DASHBOARD.**—The term "data dashboard" means any secure webpage that describes all types of data that the operator has collected on the student and how each type of data is used.

(6) **TARGETED ADVERTISING.**—The term "targeted advertising" means advertising that is presented to a student and selected on the basis of information obtained or inferred over time from such student's online behavior, use of applications, or sharing of student personal information. "Targeted advertising" does not include advertising—

(A) That is presented to a student at an online location—

a) on the basis of such student's online behavior, use of applications, or sharing of student personal information during his current visit to that online location; or

b) in response to that student's request for information or feedback; and

(B) For which a student's online activities or requests are not retained over time for the purpose of subsequent advertising.

SEC. 3. PROTECTING STUDENT PRIVACY.

(a) **IN GENERAL.**—Nothing in this Act prohibits an operator from—

(1) disclosing covered information of a student for legitimate research purposes—

(A) as required by state or federal law and subject to the restrictions under applicable state and federal law; or

(B) as allowed by state or federal law and under the direction of a school, school district, or state department of education, if no covered information is used for any purpose in furtherance of advertising or to amass a profile on the student for purposes other than educational purposes.

(b) **PROHIBITED PRACTICES.**—An operator may not knowingly—

(1) engage in targeted advertising on the operator's site, service, or application, or target advertising on any other site, service, or application when the targeting of the advertising is based upon any information, including covered information and persistent unique identifiers, that the operator has acquired because of the use of that operator's site, service, or application;

(2) use information, including persistent unique identifiers, created or gathered by the operator's site, service, or application, to amass a profile about a student except in furtherance of educational purposes; and

(3) sell a student's information, including covered information.

(c) **REQUIREMENTS.**—An operator shall—

(1) provide each student with a "data dashboard," such that—

(A) Data that is directory information or similarly discrete should be disclosed in detail to the student on the secure dashboard, whereas other data types may be described at a higher level.

(B) This section does not limit the ability of an operator to request a student's permission for collecting or sharing new types of data, subject to the following conditions: (a) the student must opt into the new type of data collection or usage on the dashboard, and (b) the data collection or usage must be for well-defined educational purposes and not targeted advertising or any other non-educational purpose.

(2) provide students with a well-defined process for requesting the deletion of their accounts and all associated data collected or produced during the account's lifetime, and within 30 days of the request, the operator shall delete the aforementioned account and data, excluding any research data that was disclosed to researchers under subsection (a) before the student made the initial request.

(3) provide students with a well-defined process for amending any data associated with them that is inaccurate or misleading, including, but not limited to, data displayed on the operator's data dashboard.

(4) provide students with the option to download, or otherwise export, all data pertaining to them, as long as—

(A) the data is already being provided to researchers;

(B) the data is directory information; or

(C) the data is covered information that can readily be exported without serving an undue burden on the operator.

(5) publish its data pipeline to students, including but not limited to—

(A) what types of data are shared with personnel;

(B) the number of personnel at each access level; and

(C) training or qualification required and received at each access level, including current employment status or other criteria for terminating access.

(6) implement and publish security measures to ensure that—

(A) only authorized individuals may access particular types of data, where authorization is established as in (5)(A-C); and

(B) an individual is immediately denied access to data when authorization is withdrawn.

SEC. 3. IMPLEMENTATION AND ENFORCEMENT.

(a) **ENFORCEMENT BY ATTORNEY GENERAL.**—Any operator determined to have violated a provision from sections (1) through (9) shall be liable for a civil penalty not to exceed one hundred thousand dollars ($100,000) per violation. The penalty shall be imposed by a civil enforcement action brought by the Attorney General or the Federal Trade Commission in the district court for the district in which the violation occurred. The penalty may be reduced at the discretion of the court if the operator can demonstrate that

(1) the violation caused no substantial harm to students; and

(2) efforts to comply were promptly taken upon notice of the violation.

Any students affected by the violation shall be notified in a timely manner.

(b) **PRIVATE RIGHT OF ACTION.**—Any student injured or adversely affected by an operator's violation of a provision from sections (1) through (9) may request both an action for injunctive relief and an action for damages in the district court for the district in which the student is located or in which the operator is located. In the case of a successful action, the operator is liable to the plaintiff for the plaintiff's damages and cost of action, including attorney fees, as determined by the court.