“Data Mining Your Children”

By Stephanie Simon

Politico

May 15, 2014

The NSA has nothing on the ed tech startup known as Knewton.

The data analytics firm has peered into the brains of more than 4 million students across the country. By monitoring every mouse click, every keystroke, every split-second hesitation as children work through digital textbooks, Knewton is able to find out not just what individual kids know, but how they think. It can tell who has trouble focusing on science before lunch — and who will struggle with fractions next Thursday.

Even as Congress moves to rein in the National Security Agency, private-sector data mining has galloped forward — perhaps nowhere faster than in education. Both Republicans and Democrats have embraced the practice. And the Obama administration has encouraged it, even relaxing federal privacy law to allow school districts to share student data more widely.

The goal is to identify potential problems early and to help kids surmount them. But the data revolution has also put heaps of intimate information about school children in the hands of private companies — where it is highly vulnerable to being shared, sold or mined for profit.

A POLITICO examination of hundreds of pages of privacy policies, terms of service and district contracts — as well as interviews with dozens of industry and legal experts — finds gaping holes in the protection of children’s privacy.

The amount of data being collected is staggering. Ed tech companies of all sizes, from basement startups to global conglomerates, have jumped into the game. The most adept are scooping up as many as 10 million unique data points on each child, each day. That’s orders of magnitude more data than Netflix or Facebook or even Google collect on their users.

Students are tracked as they play online games, watch videos, read books, take quizzes and run laps in physical education. The monitoring continues as they work on assignments from home, with companies logging children’s locations, homework schedules, Web browsing habits and, of course, their academic progress.

A report by McKinsey & Co. last year found that expanding the use of data in K-12 schools and colleges could drive at least $300 billion a year in added economic growth in the U.S. by improving instruction and making education more efficient.

Parents, however, are growing increasingly wary — and deeply frustrated. They’re finding that it’s nearly impossible to find out which companies are collecting data on their children, much less how it’s being used.

School administrators are often in the dark, too. They don’t know which digital tools individual teachers are using in the classroom. And when they try to ask pointed questions of the ed tech companies they work with directly, they don’t always get clear answers.
“When you really start digging in... they start getting antsy. It’s ‘Why are you asking this?’” said Lenny Schad, chief information technology officer for the Houston Independent School District.

“This is a problem we can’t ignore,” Schad said. It is, he said, “the wild, wild West.”

Knewton CEO Jose Ferreira finds such concerns overblown. When parents protest that they don’t want their children data-mined, Ferreira wishes he could ask them why: Is it simply that they don’t want a for-profit company to map their kids’ minds? If not, why not? “They’d rather the NSA have it?” he asked. “What, you trust the government?”

Ferreira said he often hears parents angrily declaring that their children cannot be reduced to data points. “That’s not an argument,” Ferreira said. “I’m not calling your child a bundle of data. I’m just helping her learn.”

LOOPHOLES IN AN OLD LAW

The U.S. Department of Education has called safeguarding children’s privacy a priority. “That has to be first, that has to be foremost, that’s absolutely paramount,” Education Secretary Arne Duncan said in a recent video chat posted by the department.

Yet the Family Educational Rights and Privacy Act, written when the floppy disk was just coming into vogue, offers only limited protections.

The 1974 law, known as FERPA, explicitly gives school districts the right to share students’ personal information with private companies to further educational goals.

Companies are supposed to keep standardized test scores, disciplinary history and other official student records confidential — and not use it for their own purposes. But the law did not anticipate the explosion in online learning.

Students shed streams of data about their academic progress, work habits, learning styles and personal interests as they navigate educational websites. All that data has potential commercial value: It could be used to target ads to the kids and their families, or to build profiles on them that might be of interest to employers, military recruiters or college admissions officers.

The law is silent on who owns that data. But Kathleen Styles, the Education Department’s chief privacy officer, acknowledged in an interview that much of it is likely not protected by FERPA — and thus can be commercialized by the companies that hold it.

Districts could write privacy protections into their contracts with ed tech companies. But few do.

A recent national study found that just 7 percent of the contracts between districts and tech companies handling student data barred the companies from selling it for profit.

Few contracts required the companies to delete sensitive data when they were done with it. And just one in four clearly explained why the company needed personal student information in the first place, according to the study, conducted by the Center on Law and Information Policy at Fordham University.
“We don’t know what these companies are doing with our children’s data,” said Joel Reidenberg, the Fordham law professor who conducted the study.

A White House report on big data released earlier this month recognized the risk, and called for updating FERPA. Sen. Ed Markey (D-Mass.) and Sen. Orrin Hatch (R-Utah) on Wednesday began circulating a draft bill to do just that. Their bill would tighten controls on student records and give parents the right to review — and correct — some of the information that private companies hold on their children. But the bill only covers official student educational records, not the streams of “metadata” that companies collect when kids work online.

There’s no conclusive proof any company has exploited either metadata or official student records. But privacy experts say it’s almost impossible to tell. The marketplace in personal data is shadowy and its impact on any one individual can be subtle: Who can say for sure if they’re being bombarded with a certain ad or rebuffed by a particular employer because their personal profile has been mined and sold?

Ed tech insiders will not name bad actors in their industry. But they will say this: It’s quite possible to exploit student data — and there can be a great deal of pressure to do so, especially for startups that are giving away their product for free in hopes of gaining a toe-hold in classrooms.

Unless your product is good enough to sell, “there’s this huge temptation to just make money by selling or exploiting data,” said Matthew Rubinstein, the founder and CEO of LiveSchool, which markets software that helps schools track student behavior.

Children’s personal information “is splintering across the Internet,” said Cameron Evans, Microsoft’s chief technology officer. “Anonymity is going to be more valuable than gold in the near future.”

**STUDENT RECORDS AT RISK**

Ed-tech companies divide into two main camps. Some serve as digital file cabinets for pre-existing student records; they’re basically organizational tools. Others deliver lessons and quizzes online and collect fresh data directly from students as they work.

The POLITICO examination found that both can carry privacy risks.

Take LearnBoost, a startup backed by prominent venture capital firms. It’s marketed as a “free and amazing” tool that lets teachers upload their notes on student attendance, test scores, behavior and more to a digital grade book. Any teacher can sign up, even if her district doesn’t participate.

A key element of the pitch: LearnBoost makes it easy for teachers to email the grade book to parents, students and others “as they see fit.”

LearnBoost does note in passing that confidential student data should be shared “very carefully.” But it offers no guidelines. And privacy advocates find it alarming that a for-profit startup is holding student records and making it easy for teachers to send them zipping around the Internet without supervision from the district.

The company did not return emails seeking comment.
Other sites receive huge amounts of student information directly from schools or districts. The data management site LearnSprout, for instance, stores information such as attendance records, which can be granular to the point of noting head lice, a cold, a doctor’s appointment or bereavement — to name just a few of the categories. Interactive Health Technologies stores multi-year fitness records on students, based on data from heart monitors they wear in P.E., and integrates them with “unlimited data points” from the classroom, including behavioral and nutrition records.

Knowing so much personal data is in a private company’s hands worries some parents, especially in the wake of the cyberattack that stole credit card numbers from tens of millions of Target customers last winter.

K-12 districts and contractors haven’t reported any major data breaches, but it’s been a recurring problem for colleges. In one of the worst incidents, hackers attacked the University of Maryland in February and scooped up records — including social security numbers — for nearly 300,000 students, faculty and staff.

Other companies hold more even more intimate, and potentially more valuable, information on children.

Consider the popular nonprofit tutorial service Khan Academy. It’s free. But users do pay a price: In effect, they trade their data for the tutoring.

“Data is the real asset,” founder Sal Khan told an academic conference last fall.

The site tracks the academic progress of students 13 and older as they work through online lessons in math, science and other subjects. It also logs their location when they sign in and monitors their Web browsing habits. And it reserves the right to seek out personal details about users from other sources, as well, potentially building rich profiles of their interests and connections.

After POLITICO inquired about Khan Academy’s privacy policy, which gave it the right to draw on students’ personal information to send them customized advertising, the policy was completely rewritten. The new text, posted online late last week, emphasizes Khan Academy’s commitment to protecting privacy and deletes the line about targeted advertising.

But the revised policy makes clear that Khan Academy still allows third parties, such as YouTube and Google, to place the tiny text files known as “cookies” on students’ computers to collect and store information about their Web usage. Khan Academy also states that it may share personal information with app developers and other external partners, with students’ consent.

A spokeswoman for the site said Khan Academy’s main goal in collecting data is to “help students learn effectively and efficiently.”

**MURKY PRIVACY POLICIES — OR NONE AT ALL**

Parents and teachers typically turn to companies’ privacy policies to try to figure out what student data is being collected and how it could be used. Clarity is a rarity.
Even companies that assert they do not sell personal information typically reserve the right to change that policy at any time. Most won’t notify users in the event of such a change. Instead, they recommend reading the online privacy policy regularly to see if it’s been updated.

Most policies also indicate that student information will trade hands, and may be subject to an entirely new privacy policy, if the company is sold — a common fate for a start-up.

Then there’s the legal jargon and fuzzy terminology to unravel.

**Moodle**, which many schools use as a forum for students to post work and communicate with teachers, states that it won’t share users’ personal information — “but it may be accessible to those volunteers and staff who administer the site and infrastructure.” Who are those volunteers? Are they trained to protect user privacy? The site lists an email address for users to get more information, but questions sent to that address bounced back.

Google’s privacy policy is considerably more detailed, but until recently, it did not make clear that the company scanned all emails sent through its Google Apps for Education platform, which is used by millions of students and teachers. The automated scan picked out key words that might suggest a user was, say, planning a camping trip. Google could then use that information to target ads to that individual. It did not routinely send ads to students, but it did direct them to alumni who used the Google Apps for Education platform.

After angry students filed a lawsuit, Google updated its terms of service to acknowledge the email scanning — and then announced late last month that it would stop the practice altogether for customers using Apps for Education.

Other companies don’t make any privacy policy at all available for parents to review, POLITICO has found.

The data storage and analytics firm eScholar, which sells software to help districts manage records on 20 million students — and stores some of that data on its servers — does not have a posted privacy policy. Spokeswoman Ann Tarasena said the company is working on it. In the meantime, eScholar writes privacy protections into its contracts with districts. It wouldn’t release the contracts — citing privacy concerns.

On Thursday, responding to questions raised by this article, the company posted online a statement of its **general privacy principles**, including a pledge not to sell student data.

Then there’s **Panorama Education**, a data analytics platform used by thousands of schools and backed by investors including Facebook’s Mark Zuckerberg and actor Ashton Kutcher.

CEO Aaron Feuer said the company abides by each district’s privacy rules, but it does not have a blanket policy to share with the public.

The lack of consistent standards troubles Sen. Markey, who has become a leading voice on consumer privacy in Congress.
“The goal here should be to help scholars make the grade,” Markey said, “not help companies make a sale.”

DATA DEMANDS ESCALATE

In recent months, more than 30 public school districts from Bainbridge Island, Washington, to Broward County, Florida, have signed partnerships with a nonprofit called Code.org. The organization gives schools free curricular materials and teacher training to set up computer science classes.

All it asks for in exchange: Data. Lots and lots of data.

Code.org requires that its partner schools turn over up to a dozen years of academic records, including test scores, on every participating student, according to a model contract reviewed by POLITICO.

In addition to their official academic records, Code.org collects huge amounts of new information on participating students as they watch the tutorials and do the activities on its website. It collects their computer login, email address and password and captures their interaction with the website, including searches conducted on the site.

Code.org’s privacy policy is a bit confusing. It says it does not share personal information with third parties. It promises not to share it, either, with the organization’s donors or sponsors, among them Facebook’s Zuckerberg and Microsoft’s Bill Gates.

But the policy goes on to say it may provide personal information to “schools, teachers and affiliated organizations.” It explicitly states that Code.org does not control how that information “is later used by them or shared with others.”

The policy doesn’t define “affiliated organizations” or explain how access is determined. Nor does it explain what Code.org does with its voluminous student files or how it protects them.

Spokeswoman Roxanne Emadi declined to discuss those issues.

Officials in some districts that have signed up with Code.org said they were comfortable with handing over the data because they assumed it would be aggregated and anonymized — though the contract makes no mention of that — and used to gauge the effectiveness of the program. “That kind of analysis and research goes on all the time,” said Robert Runcie, superintendent in Broward County. “It’s not a problem.”

Others said they didn’t realize when they signed the contract how much data would be turned over.

On Thursday, in response to this article, Code.org CEO Hadi Partovi posted an open letter stating that the organization only sought anonymized data and would update and clarify its privacy policy.

To Doug Levin, who runs an association of state educational technology directors, the ambiguity of the Code.org policy is an astonishing example of how little attention is being paid to protecting student privacy.
“That’s just unacceptable,” he said. “I mean, you’re just throwing potentially sensitive information over the wall in the hopes that there won’t be any issue.” The contracts highlight the danger of the ed tech explosion, Levin said: When it comes to protecting privacy, “the rules of the road are not real clear.”

**THE PROMISE OF BIG DATA**

For all the concerns about privacy, education reformers are adamant that the digital revolution must be allowed to flourish.

Already, publishers are producing digital textbooks that can effectively read students’ minds, figuring out when they’re on the verge of forgetting key concepts and sending them text, video or quizzes to fix the facts firmly in their memory.

Even more intimate tracking may be possible in the future: The Bill & Melinda Gates Foundation funded a **$1.4 million research project** in 2012 to outfit middle-school students with biometric sensors designed to detect how they responded on an a subconscious level to each minute of each lesson. The results suggested the sensors could be useful for teachers, foundation spokeswoman Deborah Robinson said.

“We’re really just at the beginning of truly leveraging the power of data to transform the process of teaching and learning,” said Aimee Rogstad Guidera, executive director of the **Data Quality Campaign**, which urges states to develop responsible policies for data-driven education.

“When we take the time to explain to parents why this is good and how it’s going to help, they’re fine with it,” Guidera said.

Not all parents, however, are convinced.

To Barmak Nassirian, a father of two and grassroots privacy activist, the question boils down to this: No matter how well they safeguard the data, no matter how stringent their privacy policies, do you want private companies “to get into your kid’s head and mine the learning process for profit?”

Investors, after all, are pouring into the sector because they expect it to make money, not because — or at least, not only because — they believe it will help kids learn.

“Our mission isn’t a social mission,” said Michael Moe, co-founder of GSV Capital, a leading investor in ed-tech companies. “They’re there to create return.”

More than $650 million flowed into technology firms serving the K-12 and higher education market last year. That’s nearly double the $331 million invested in those spheres in 2009, Moe said. Nationwide, the market for education software and digital content stands at nearly **$8 billion**, according to the Software & Information Industry Association.

It’s not entirely clear that all those apps boost achievement; a recent national survey commissioned by the Gates Foundation found just 54 percent of teachers considered the digital tools their students use frequently to be effective.

Given that uncertainty, the data companies collect on students could be their most valuable assets.
Publishers of digital textbooks, for instance, could potentially use their insights on students’ academic progress to pitch them — or, more likely, their parents — new products targeted directly at their needs.

“In the industry, there’s a lot of desire to do that,” said Andrew Bloom, chief privacy officer for McGraw-Hill Education. Bloom stressed that McGraw-Hill has no plans to do such marketing unless school districts consent.

Khaliah Barnes, director of the student privacy project for the Electronic Privacy Information Center, can imagine another scenario: Companies with rich student dossiers could market aptitude and attitude profiles to college admissions or corporate recruiting offices.

“As an employer, that’s the sort of profile I would want to buy: Who can solve a problem quickly? Who has the tenacity to finish all the problems? Who drops off quickly?” Barnes said.

Ferreira, the CEO of the New York data analytics firm Knewton, said he’s not planning to create such profiles. “But I suppose I can imagine a future where it happens,” he said. “I’m not sure how I feel about that.” If such profiles were to come into use, he said, Knewton would not sell or share them without students’ consent.

A model state bill drafted by the American Legislative Exchange Council, a conservative lobbying group, could make such targeting more likely; it would set up a central state database for student records and allow colleges or businesses to browse them in search of potential recruits.

Companies might also seek to mine student profiles to find customers uniquely vulnerable to their sales pitches. For instance, young adults who struggled with high-school math could be bombarded with ads for high-priced payday loans, Barnes said.

Such prospects may sound far-fetched, but the recent White House report on big data acknowledged it as a very real possibility. Data collected on children as they take advantage of educational services “could be used to build an invasive consumer profile of them once they become adults,” the report concluded.

Knewton’s Ferreira is impatient with alarmist scenarios and anxious parents.

He once described education as “the world’s most data-mineable industry, by far” — and he has raised $105 million from investors who share that vision. By next year, he expects to be mapping the minds of 10 million students. If he can identify who among them will struggle with fractions next Thursday, he can also recommend resources to help them before they hit that wall.

Ferreira has a tough time understanding how anyone could object to data mining when it has such power.

“It just helps children,” Ferreira said. “That’s all it does.”