

Big Data Transforms Education Research

Can machine learning unlock the keys to great teaching?

by MICHAEL J. PETRILLI

FOR DECADES, education technophiles have envisioned a future wherein gee-whiz devices and engaging digital applications whisk students away from the doldrums of traditional classroom instruction and into a fun world of beeping computers, self-paced lessons, and cloud-based collaboration.

That may yet come to pass—and at some outlier schools, is already here—but don't be surprised if the true transformative power of education technology is most evident when it comes to something old-fashioned: basic education research. The declining cost and easy availability of substantial computing power may enable us finally to unlock the black box of the classroom, giving scholars and teachers much more insight into what is and isn't working. Technology can do more than just keep students engaged; it can equip teachers, school and district leaders, and policymakers with the sort of insights and analytics that can help them make better decisions for students.

A Challenging Research Subject

Studying the actual behavior of teachers and students has always been a difficult and expensive proposition. The most respected approach involves putting lots of trained observers—often graduate students—in the back of classrooms. There, they typically watch closely and code various aspects of teaching and learning, or collect video, take it back to the lab, and spend innumerable hours coding it by hand.

This kind of methodology has helped the field gain significant insights, such as the importance of teachers asking open-ended questions, and how better to evaluate teachers' practice, à la the Gates-funded Measures of Effective Teaching initiative. But it's incredibly labor-intensive, costs gobs of money, and thus may not be practical.

Alternatives to observational studies are much less satisfying. The most common approach is to survey teachers about their classroom practices or curricula, as is done with the background questionnaires given to teachers as part of the National



Assessment of Educational Progress (NAEP). Though useful, these types of surveys have big limitations, as they rely on teachers to be honest and accurate reporters of their own practice—which is tough even with positive intentions. It's not easy to remember what you taught months ago, and teachers might also try to tell researchers what they think they want to hear or choose responses that cast themselves in a positive light. Another approach, asking teachers to keep logs detailing their work, such as how they spend time, is somewhat more reliable, but

still far from perfect. It is also time-consuming, thus stealing precious minutes and hours from teachers' most important work: helping students learn.

Not surprisingly, the research base on the real stuff of education—instructional practices, homework assignments, the curriculum as it is actually taught—is remarkably thin. Scholars have found easier, cheaper, and more fruitful yields from mining administrative data sets, usually stemming from compliance reports at the school or district level, than from collecting detailed information about what's happening in real classrooms in real time. This has left the field, and policymakers, with a huge blind spot about what teachers and kids are doing, and what might or might not be working.

“Machine Learning” to Track Student Learning

Enter the machines. What if we didn't need to have graduate students crouching in the back of classrooms in order to catalog the play-by-play of classroom instruction? What if, instead, we could capture the action with a video camera or, better yet from a privacy perspective, a microphone? And what if we could gather that information not just for an hour or two, but all day, 180 days a year, in a big national sample of schools? And what if we could then use the magic of machine learning to have a computer figure out what the reams of data all mean?

This possibility is much closer than you might imagine, thanks to a group of professors who are teaching computers to capture and code classroom activities. Martin Nystrand

(University of Wisconsin-Madison), Sidney D’Mello (University of Notre Dame), Sean Kelly (University of Pittsburgh), and Andrew Olney (University of Memphis) are interested in helping teachers learn how to ask better questions, as research has long demonstrated that high-quality questioning can lead to better engagement and higher student achievement. They also want to show teachers examples of good and bad questions. But putting live humans in hundreds of classrooms, watching lessons unfold while coding teachers’ questions and students’ responses, would be prohibitively costly in both time and money.

So with funding from the Institute of Educational Sciences, this team of researchers decided to teach a computer how to do the coding itself. They start by capturing high-quality audio with a noise-canceling wireless headset microphone worn by the teacher. Another mike is propped on the teacher’s desk or blackboard, where it records students’ speech, plus ambient noise of the classroom. They take the audio files and run them through several speech-recognition programs, producing a transcript. Then their algorithm goes to work, looking at both the transcript and the audio files (which have markers for intonation, tempo, and more) to match codes provided by human observers.

The computer program has gotten quite good at detecting different types of activities—lectures vs. group discussion vs. seatwork, for example—and is starting to be able to also differentiate between good questions and bad. To be sure, D’Mello told me, humans are still more reliable coders, especially for ambiguous cases. But the computers are getting better and better, and good enough that, with sufficient data, they can already produce some very reliable findings at a fraction of the cost of a people-powered study.

It’s even easier, of course, if the underlying instructional data are digital to begin with. That’s the specialty of Ryan Baker, associate professor of teaching, learning, and leadership at the University of Pennsylvania. He and his team examine the “digital traces” of students’ interactions with digital applications—their key strokes, pauses, and answers when working on online math programs, for example. They then build algorithms to make sense of them. Their research starts by asking humans to watch students at work; their insights are fed into their computer models, which learn to replicate the human coding with enough time and data.

Such research has already borne fruit. Baker’s team and its computer have shown that more students become bored, then disengaged, when the material is too hard than when it is too easy. Short periods of confusion and frustration are good; long periods indicate that the student has given up. And some “off task” time—as long as a minute or two—is OK, as students tend to come back refreshed and ready to tackle whatever they are working on. Thus,

teachers should allow kids some breathing room rather than cracking the whip the second they see students get distracted.

Putting Data to Work

This is incredibly useful information, the kind that can help teachers improve their practice and boost the efficiency and effectiveness of students’ time in class. Imagine if such studies—both of traditional classroom practices and the digital variety—became much more common. Large national studies like NAEP could complement teacher surveys with the collection of audio, every day, all day, in a big sample of schools. Plus, they could capture the digital activity of students, and ask teachers to scan student assignments and tests so those could be analyzed as well.

We would finally have an accurate picture of what’s actually being taught in U.S. schools. And if we combine that with state administrative and achievement data, and put it in the hands of competent analysts, we’d have a better way to examine which teacher practices, curricula, use of time, and on and on, are related to improved student learning. We could see whether teachers whose students make the largest gains really do make greater use of the concrete practices that Doug Lemov describes in *Teach Like a Champion*, for example. And we could determine whether and where there are equity gaps in effective teaching, the level of challenge of student assignments, and much else that might be addressed in order to narrow the achievement gap.

Big hurdles remain, to be sure. The biggest aren’t technological, but political: Chronicling classrooms in minute detail will not go over well with all teachers, even if researchers promise that the data will be used for research purposes only. Nor will privacy-minded parents be thrilled; security protocols will need to be established that give everyone involved confidence that the audio recordings won’t fall into the wrong hands. And scholars will need to be careful not to make causal claims based on data sets that aren’t subject to experimental designs; the sheer quantity of data can’t make up for the lack of controls and random assignment. Big data alone can be a boon to “hypothesis generation,” but we’ll still need traditional studies in which teachers are asked to adopt new practices to learn whether the practices work.

Still, the power duo of big data and machine learning will enable us to build a research enterprise that actually improves classroom instruction, regardless of how traditional or technology-infused the instruction might be. That’s enough to make a computer smile.

Michael J. Petrilli is president of the Thomas B. Fordham Institute and an executive editor at Education Next.

What if we could capture the play-by-play of classroom instruction not just for an hour or two, but all day, 180 days a year, in a big national sample of schools?