



Studying Teacher Moves

In July 2011, **Bill Gates** told the *Wall Street Journal*,

“I believe in innovation and that the way you get innovation is you fund research and you learn the basic facts.... I’m enough of a scientist to want to say, ‘What is it about a great teacher?’”

As a “practitioner” of sorts, I’ve wondered the same thing for 15 years. The K–12 school sector generates little empirical research of any sort. And of this small amount, most is targeted to policymakers and superintendents, and concerns such matters as the effects of class size reduction, charter school attendance, or a merit-pay program for teachers. Why is there virtually no empirical education research meant to be consumed by the nation’s 3 million teachers, answering *their* questions?

Those 3 million teachers generate about 2 billion hour-long classes per year. We do not know empirically which “teacher moves,” actions that are decided by individual teachers in their classrooms, are most effective at getting students to learn. Why doesn’t this kind of research get done?

Mr. Gates has part of the answer. Money. For 2011, the Microsoft R&D budget is \$9.6 billion, out of total revenue in the \$60 billion range. The U.S. Department of Education’s Institute of Education Sciences (IES) represents only a fraction of total education research, but its budget gives some perspective: IES spends about \$200 million on research compared to more than \$600 billion of total K–12 spending. So, 15 percent to upgrade Microsoft, 0.03 percent

A practitioner’s take on what is blocking the research teachers need

to upgrade our nation’s schools. And while Microsoft’s research is targeted to the bottom line (\$8.6 billion is on cloud computing, the profit center of the future), IES spends almost nothing examining the most important aspect of schools: the decisions and actions that individual teachers control or make.

By MICHAEL GOLDSTEIN

One IES project is the What Works Clearinghouse (WWC), established in 2002 to provide “a central and trusted source of scientific evidence for what works in education.” The WWC web site lists topic areas like beginning reading, adolescent literacy, high school math, and the like. For each topic, WWC researchers summarize and evaluate the rigor of published studies of products and interventions. One might find on the WWC site evidence on the relative effectiveness of middle-school math curricula or of strategies to encourage girls in science, for example.

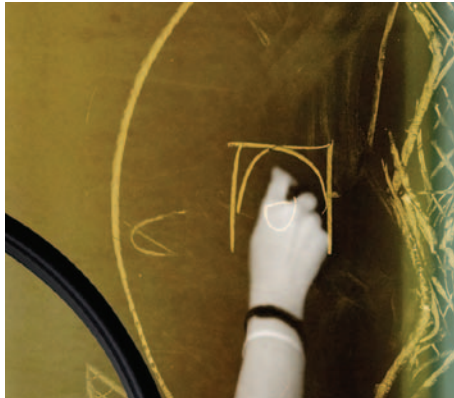
But there is almost nothing examining the thousands of moves teachers must decide on and execute every school day. Should I ask for raised hands, or cold-call? Should I give a warning or a detention? Do I require this student to attend my after-school help session, or make it optional? Should I spend 10 minutes grading each five-paragraph essay, 20 minutes, or just not pay attention to time and work on each until it “feels” done?

And the WWC’s few reviews of research on teacher moves aren’t particularly helpful. A 63-page brief on the best teaching techniques identifies precisely two with “strong evidence”: giving lots of quizzes and asking deep questions. An 87-page guide on reducing misbehavior has five areas of general advice that “research supports,” but no concrete moves for teachers to implement. It reads, “[Teachers should] consider parents, school personnel, and behavioral experts as allies who can provide new insights, strategies, and support.” What does not exist are experiments with results like this: “A randomized trial found that a home visit prior to the beginning of a school year, combined with phone calls to parents within 5 hours of an infraction, results in a 15 percent drop in the same misbehavior on the next day.” If that existed, perhaps teachers would be more amenable to proposals like home visits.

By contrast, a fair number of medical journals get delivered to my house. They’re for my wife, an oncologist. They’re practical. In each issue, she learns something along these lines: “When a patient has this type of breast cancer, I currently do X. This study suggests I should do Y.” There is a bit on medical policy, but most of the information is meant for individual doctors in their day-to-day work.

That’s not to say that we shouldn’t conduct research on education policy. My own work has certainly benefited from it. For example, the quasi-experimental study by economists Tom Kane and Josh Angrist on Boston charter schools, which compared the winners and losers of charter admission lotteries, helped

change the Massachusetts law that had blocked the creation of new charters. The change enabled me to help launch a new charter school, MATCH Community Day. My point is simply that relative to education policy research, there is very, very little rigorous research on teacher moves. Why? Gates knows it’s more than a lack of raw cash; it’s also about someone taking responsibility for this work. “Who thinks of it [empirical



Researchers care about next August (when test scores come in). Teachers care more about solving today’s problems.

research on teachers] as their business?” he asked. “The 50 states don’t think of it that way, and schools of education are not about [this type of] research.”

I agree, but I contend there are a number of other barriers. The first is a lack of demand.

The Demand Side

Why aren’t teachers clamoring for published research? One reason is that researchers generally examine the wrong dependent variable. Researchers care about next August (when test scores come in, because they can show achievement gains). Teachers care about that, too, but they care more about solving today’s problems (see sidebar, page 26).

A second issue is that researchers don’t worry about teacher time. Education researchers often put forward strategies that make teachers’ lives harder, not easier. Have you ever tried to “differentiate instruction”? When policy experts give a lecture or speak publicly, do they create five different iterations for their varied audience? Probably not.

The return on investment for teacher time and the opportunity cost of spending it one way rather than another is rarely taken into account. In what other, valuable ways could teachers be spending the time taken up with building “differentiation” into a lesson plan? They could phone parents, tutor kids after school, grade papers, or analyze data. Much research implies that teachers should spend more time doing X while not indicating where they should spend less time.

Teachers don’t trust research, and understandably so. There’s a lot of shoddy research that supports fads. Experienced

teachers remember that “this year’s method” directly contradicts the approach from three years ago. So they’d rather go it alone. Newer teachers pick up on the skepticism about research from the veterans.

Unlike medical research, teacher research rarely examines possible side effects, and whether they are short-term aggravations or can be expected to persist. Imagine that a teacher reads an article arguing that students benefit from being asked “higher-order questions.” She begins doing that. Some students, surprised at this new rigor, are frustrated. Some students throw up their hands and give up. Misbehavior ensues.

Student frustration is probably a fairly predictable short-term side effect of asking higher-order questions. If she isn’t being properly warned, a teacher might quickly abandon this technique.

For all these reasons, the 3 million teachers aren’t forming picket lines to demand research.

Do We Know What Works?

Neither policy camp, reformers nor traditionalists, care much about research into teacher moves, either. Some traditionalists see teaching as an art, one that cannot be subjugated to quantitative analysis (“every teacher is different”). Others aren’t averse to research; they simply don’t see it as a priority. They’d prefer that limited resources be used to fight poverty, not to improve students’ day-to-day classroom experiences.

Meanwhile, some reformers argue “we already know what works,” and we just need to scale it.

As part of the “reformer” community, I find this troubling. From charter opponents like Diane Ravitch to supporters like education secretary Arne Duncan, there’s agreement that “some charter schools work.” Furthermore, there’s strong evidence that the charters that succeed tend to be “No Excuses” schools. So do we know what works?

I’m the founder of one of those charter schools; our high-school students have the highest value-added gains of all 340 public high schools in Massachusetts. I’m also the founder of a

small teacher residency program that supplies teachers to schools like KIPP (Knowledge Is Power Program). Many of us would agree to a very different proposition: We know teacher moves “that work” to some extent, enough to create very large achievement gains, but we don’t know teacher moves well enough to get our college graduation rate near where we’d

like it to be. Nor do we know how to help teachers do these moves more efficiently, so that their jobs are sustainable.

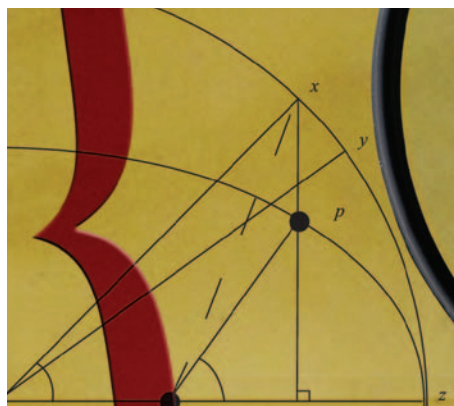
Without a massive uptick in our knowledge of teacher moves, we’ll continue on the current reform path. That path is a limited replication of No Excuses schools that rely on a very unusual labor pool (young, often work 60+ hours per week, often from top universities); the creation of many more charters that, on average, aren’t different in performance from district schools; districts adopting “lite” versions of No Excuses models while pruning small numbers of very low performing teachers; and some amount of shift to online learning. Peering into that future, I don’t see how we’ll generate a breakthrough.

Bridging the Divide

The final barrier to research on teacher moves is the divide between practitioners and researchers. My analogy is a 5th-grade dance. Boys stand on one side. Girls stand on the other. There is very little actual dancing. In this case, teachers are off to one side, and quantitatively oriented researchers are on the other.

After a while, the boys go into the hallway and talk about video games. Similarly, quantitative researchers find the transaction costs of setting up experiments are too high and give up on doing research about teacher actions. They take their problem-solving marbles and find other data sets to crunch.

Girls see that the boys aren’t around anymore. So they dance with each other. Teachers and school leaders, if they like to learn, do so through observation of and conversation



Some traditionalists see teaching as an art. Others are not averse to research; they simply don’t see it as a priority.

regarding perceived “best practices.” There aren’t many practitioners who care about rigorous empirical research.

With all these barriers, is there much hope? There’s not going to be a pot of gold in this funding environment. If research on teacher moves matters, we need to be more creative about catalyzing the low-hanging fruit. That would

What Do Teachers Want to Know?

If we're going to get researchers to dance with the teachers, it makes sense to focus on topics that teachers care about. Here are the things I think "well-intentioned teachers" care most about:

1. How to be more efficient. Many teachers want to work less without being neglectful. Or they'd like to free up time to invest in new priorities.
2. How to manage the classroom so kids behave better, thus lowering the "misbehavior tax" on learning. If a middle school teacher can "reset" the class only 3 times per period, instead of 5, that's probably 1,440 fewer times per year that he has to deal with misbehavior. (By "reset," I mean when a teacher says something like, "Guys, come on. I need your eyes on me. I need you to settle down. Joey, that means you. I'm going to wait until I have everyone's eyes.")
3. How to motivate and generate student effort, especially, how to "flip" kids who arrive having not worked hard in previous classes or years. This includes both getting kids to exert effort during class and getting them to work hard at home.
4. How to get kids to remember material that they seemingly once knew. Cognitive science has moved the ball forward here; now we need applied experiments with teachers.
5. How to best explain particular ideas and concepts. Each year, tens of thousands of math teachers try to get kids to understand the notion that division by zero does not exist.

mean identifying practitioners who are unusually interested in randomized research, and connecting them with doctoral students who are unusually interested in teachers and teaching.

What does it look like when practitioners and researchers dance together? Here is one example.

In July 2010, I asked Harvard economist Roland Fryer for some help. My research question was fairly simple: Do teacher phone calls to parents "work"?

In our school, teachers proactively phone parents. Typically, the parents have not been heavily involved in their children's previous schools. We believe that phone calls to parents help teachers generate improved decorum, effort, and ultimately learning from students. (Sometimes the calls to parents are supplemented with teacher calls to students) These parent relationships seem to be linked to very high parent-satisfaction ratings, and in turn we have thought those

were related to our high test-score growth. Truth be told, however, we just don't know whether this is a productive use of teachers' time.

Fryer enlisted two doctoral students, Shaun Dougherty and Matt Kraft, from the Quantitative Policy Analysis in Education program at the Harvard Graduate School of Education. These two did an amazing job, operating skillfully within our school to do the randomized study. From their findings:

"On average, teacher-family communication increased homework completion rates by 6 percentage points and decreased instances in which teachers had to redirect students' attention to the task at hand by 32%."

This collaboration worked for several reasons. First, we have a teacher residency embedded in our charter school, so I had 24 student teachers who could be fairly easily randomized during the summer school session. Second, a professor I trusted chose the graduate students who would

conduct the research. These guys were, in my view, dispassionate. I've tried to work before with grad students who have strong preexisting beliefs about what they'll find (typically with a "progressive" lens), and it was difficult to gain real knowledge. (Researchers often feel the same way about practitioners, that we're searching for marketing, not truth). Also, Fryer paid them a stipend; in my experience, graduate



In our school, teachers proactively phone parents. Truth be told, however, we just don't know whether this is a productive use of teachers' time.

students working for free, and only for credit of some sort, don't always follow through.

The cost of the two graduate students was not the only expense. In our experiment, at any given time, there were 16 classrooms in action. The researchers needed to hire 16 observers to carefully code student behavior for a few weeks. The total bill was around \$10,000. Kraft and Dougherty found a Harvard grant of \$1,000. The rest I needed to pay.

Once we'd designed the experiment, I needed to explain it to my team: the principals of our high school and middle school, and the student teachers who were involved. These are people I know well, and they generally trust me. Still, this buy-in phase required expending both time and "relationship capital," a resource that gets spent down and must be built back up over time. Using student teachers was also of benefit. It would have been tough to randomize our regular teachers. Their belief in the efficacy of parent communication is so strong I suspect many would have doubted the value of changing their normal routines.

There were other costs to the experiment. The head of our teacher-prep program spent many hours handling the experiment's complex logistics, including a permission slip for parent consent. He could have spent those hours coaching these student teachers, which is the main task I was paying him to do.

All of these issues reflect transaction costs: finding the right people and then doing the right study well takes time, effort, and money.

Researching Teacher Moves

Think of the Human Genome Project. When the project started, scientists didn't know how many genes there were; now they believe the number is 20,000 to 25,000.

We don't know how many teacher moves there are. The number is certainly high but not infinite, maybe 200, 2,000, nobody knows. Presumably, there are some unusually high-yield teacher moves across all contexts, some moves that are high yield but only in specific situations or contexts, and other less powerful moves. There is undoubtedly lots of interaction effect among many moves. Mapping all of this might be called the Teaching Move Genome Project, and at the beginning it would be a scary undertaking.

Absent this work, what do we have? Perceived best practices, often buttressed by observation or nonrandomized studies. In his best-selling book *Teach Like a Champion*, Doug Lemov describes 49 teaching moves he has observed in the nation's top charter schools. At the University of Michigan, Deborah Ball and her colleagues are close to unveiling



Imagine a group of trials that would ask the questions, Do all of the moves work? Are any particularly successful?

a list of 88 math teacher moves. Lee Canter's *Assertive Discipline* and Jon Saphier's *Skillful Teacher* discuss scores of moves, like the "10-2" rule (have kids summarize for 2 minutes in small groups after 10 minutes of teacher-led instruction), much of it supported by nonrandomized research. On the basis of its observations of effective teachers, Teach For America (TFA) promotes 6 teacher behaviors and 28 component parts, like "plan purposefully" or "set big goals"; none are specific moves.

What would a series of randomized trials look like? Let's apply it to Lemov's 49. Imagine a group of trials that would ask the questions, Do all of the moves work? Are any particularly successful? How does the degree of teacher buy-in interact with effectiveness? What are the "costs" of these moves?

An example from Lemov is "Right Is Right." The idea is that when a kid gives an answer that is mostly right, the teacher should hold out until it's 100 percent correct. Lemov describes various tactics the teacher can use to elicit the 100 percent right answer from the student (or first from another student, before having the original student repeat or extend the correct answer).

The obvious cost of implementing this move is time. These back-and-forths add up to lost minutes each period when other topics are not being discussed. A less skillful teacher might be drawn into a protracted discussion, when her next best alternative (simply announce the 100 percent right answer, and move on) might work better. We just don't know.

Back in 2003, education researchers David Cohen, Stephen Raudenbush, and Deborah Ball argued that "one could make accurate causal inference about instructional effects

only by reconceiving and then redesigning instruction as a regime, or system, and comparing it with different systems.” That suggests “a narrower role for survey research than has recently been the case in education, and a larger role for experimental and quasi-experimental research. But if such studies offer a better grip on causality, they are more difficult to design, instrument, and carry out, and more costly.”

Still, we need a better grip on causality. So who would undertake this cost?

A Proposal

Once again borrowing some terminology from medicine, I propose a typology of trials, delineating phases in a continuum.

Phase 1 trials would be small, nongeneralizable empirical studies of teacher moves. These could be randomized, single-subject, or regression discontinuity, but the dependent variable would not be year-end test scores. Instead, we’d look for next-day or next-week outcomes: measurable effects on student behavior, effort, or short-term learning.

Who would decide what moves to test? Some would be proposed by established authors and thinkers in the teaching field. Some would come from the nation’s 3 million schoolteachers, possibly with crowd sourcing to identify the most-promising ideas. Some would come from academic researchers, particularly those from other fields, like psychology, who may offer unusual insights. But for the next level, testing competing ideas, I’d suggest we draw heavily on teacher opinion, particularly a group of teachers selected

for their stated willingness to try new methods (if they are supported by research).

Phase 2 trials would test promising teacher practice from Phase 1 on a larger, more varied teacher pool to see if the next-day outcomes held up, probably across different types of schools. Again, the dependent variable is short-term student response.

Phase 3 trials would be randomized trials in which teachers combine multiple moves that emerge from Phase 2. In the end, our bottom line is student learning, and Phase 3 trials are combinations of moves that are measured to see if they bolster year-end student learning gains.

Medical researchers have found that treating some illnesses requires a drug “cocktail,” that is, no one medicine by itself works as well as the combination of several. The same approach might work in education: it could be that individual teacher moves by themselves cannot create measurable year-end achievement gains in students, but combining many together can.

My proposal is that each of the nation’s 1,200-plus schools of education and teacher prep programs conduct one randomized trial on a teacher move each year: Phase 1, Phase 2, or Phase 3. They’d do that by recruiting alumni into a network of experienced teachers willing to participate. The advantage is that once you pay the one-time transaction costs of finding these teachers, the ongoing expenditures related to persuading them to participate, and securing permission from families and principals, decline.

Once that network existed, it would function like a laboratory. Various Phase 1 experiments could be run through it, with small numbers of teachers at first, so that many experiments could be run concurrently. Larger numbers of teachers would be included in more promising Phase 2 validation experiments. Of course, there would be selection bias in terms of which teachers are willing to be participate in this sort of work, and other imperfections. But in the end, experiments could build on proven results from previous ones. Multiple ed schools would combine their networks for Phase 3 trials.

By itself, no single experiment would be that important. Instead, it would be like cancer research: thousands of people each trying to answer small questions in a very rigorous way...which would add up to promising treatments.

The goal is an affordable system for conducting teacher research that teachers would actually consume, that would address both the implementation challenges and the high transaction costs for researchers and practitioners in creating such research. Until that exists, I’ll see you at the 5th-grade dance.

Michael Goldstein is the founder of MATCH Charter School and MATCH Teacher Residency, in Boston.



“Ah, Miss Brimsley, I ask you: Which one of us has truly failed?”