

March 2021

The narrow path to do it right:

Lessons from vaccine making for high-dosage tutoring

By Michael Goldstein and
Bowen Paulle

THOMAS B.
FORDHAM
INSTITUTE
ADVANCING EDUCATIONAL EXCELLENCE

CONTENTS

Introduction	1
I. It's weird in there: Virus cells and tutorials	4
II. Cells and schools constantly create <i>new</i> problems	6
III. A respectful disagreement about research	8
IV. How to move forward	11
Coda	14
About the authors	16

Introduction

High-dosage tutoring (HDT) is “having a moment”—in the USA, Netherlands, and UK. See [here](#), [here](#), [here](#), [here](#) (UK), [here](#), and here at [Fordham](#).

Research-validated success stories have piled up over the last decade. Now “Covid-19 learning loss” has poured gasoline on the HDT fire. Problem, meet solution?

Andrew Rotherham, however, weighs in with a cautionary note. He writes:

If you invest in the silver-bullet market, there is a buy opportunity coming in tutoring. Not just any tutoring, high-dosage tutoring. The word itself sounds exciting—high-dosage!

Here is how these things tend to go. New idea—or not new but reintroduced idea—widely implemented through a funding and think piece gold rush. And widely implemented in uneven ways with little fidelity to the research because of the haste and good intentions coupled with lack of capacity around the field. End result: Good idea gets discredited because, on average, it shows little if any impact. You see this around the ed tech sector, class size reduction, teacher evaluations, some reading initiatives, and charter schools.

Rotherham nails it.

This white paper is motivated by our shared fear that HDT will scale up badly, and our hope that this unfortunate outcome can be avoided.

Some quick background: We met on a basketball court thirty years ago. We’ve chased two education reforms in our careers: school culture and high-dosage tutoring. We each penned books about school culture, with Mike outselling Bo eight copies to six copies (two more cousins). No traction there.

When it comes to high-dosage tutoring, however, we’ve been part of some successes, in the USA and in the Netherlands.

Mike’s effort in Boston—[Match Education](#)—achieved large gains. Harvard’s Roland Fryer led a replication of HDT in [Houston](#). Alan Safran of Saga Education improved and scaled the program in [Chicago](#), New York City, and beyond. Mike Duffy and Jared Taillefer turned the idea into [Great Oaks](#), a HDT-fueled charter school network in cities like Wilmington, [Newark](#), and Bridgeport.

[The Amsterdam program](#), where Bo was an advisor, had students, after half a year, making math gains between 0.43 and 0.70 standard deviations—huge gains. The Bridge Learning Interventions’ version of HDT has been replicated in multiple Dutch cities, and is now scaling up.

We are convinced high-dosage tutoring can work. Often does work. HDT has huge potential. But since every other education idea that's been scaled has failed, is there *any* way to avoid that fate here?

Hell if we know! We're terrible at politics.

Our contribution is an effort at an explainer on why HDT will *not* respond well to a "just put money behind it" policy effort.

[As John Arnold says](#), "How to get the benefits of small-scale instruction at scale is one of the most important questions to be addressed." When asked whether a lack of willingness or investment was the problem, Arnold said, "There have been lots of small-scale programs. The evidence is clear. The problem is scaling."

And to explain the tutor scale-up problem, we turn to an apropos analogy: vaccine development.

"High-dosage tutoring" has become the common parlance. But we mean "[high-impact tutoring](#)," meaning that, irrespective of dosage, kids actually made large learning gains when measured in a randomized control trial. It's possible to get high impact with high or low dosage. But it's easier with higher doses.

Susanna Loeb's new best-practices effort at the Annenberg Institute [uses this language](#):

Tutoring is a form of teaching, one-on-one or in a small group, towards a specific goal.

High-impact tutoring leads to substantial learning gains for students by supplementing (but not replacing) students' classroom experiences. High-impact tutoring responds to individual needs and complements students' existing curriculum.

We recognize high-impact tutoring programs as those that either have directly demonstrated significant gains in student learning through state-of-the-art research studies or have characteristics that have proven to accelerate student learning.

These characteristics of high-impact tutoring programs currently include: substantial time each week of required tutoring, sustained and strong relationships between students and their tutors, close monitoring of student knowledge and skills, alignment with school curriculum, and oversight of tutors to assure quality interactions.

“There have been lots of small-scale [tutoring] programs.... The problem is scaling.”

Other “personalized learning” options exist and may promote student learning by replacing traditional class periods, but we do not include them under our umbrella definition of tutoring. For example, at this point in time we are not focusing on initiatives such as: pull-out services (e.g., when students receive personalized help instead of attending a class), in-class small group instruction by a second teacher (e.g., co-teaching), or learning pods.

Mike already co-wrote a [cautionary essay](#) for the Brookings Institution on this very topic. But the final product wasn't cranky enough for his taste. That's because his dear friend and co-writer, economist Matt Kraft, is much more [optimistic](#) about scaling HDT.

So, Mike turned to Bo, who shares Mike's (and Rotherham's) glass-half-empty views. Together, we'll explain why HDT is quite hard to scale, and describe a narrow path to doing it right.

I. It's weird in there: Virus cells and tutorials

[Derek Lowe](#) writes about vaccine development:

It's not easy—especially when you're a mere chemist—to picture what's really going on inside a cell.

The sorts of pictures that most of us tend to use (two blobs to represent a ribosome, little snaky line curving out from it to represent a new protein) are helpful memory devices, but have very little to do with reality

...For example, the inside of the cell is constantly experiencing the formation (and dissolution) of all sorts of tiny non-mixing liquid droplets of concentrated proteins and RNA species, like a bottle of salad dressing a few seconds after you shake it up, but without the further separation into two bulk phases.

We'll guess that readers agree with Derek. Folks like us—decidedly not even “mere” chemists—can't conjure what's really going on inside a cell.

However, we're educators! We *think* we know what's going on inside tutorial sessions.

Nope. Unless you've been actually observing hundreds of real-life sessions, you're like Derek. You have a rough idea that's probably off target. The pictures that most of us use, an adult explaining something to a kid, are helpful memory devices, but they often have very little to do with reality.

Did you picture this?:

I would try to ask the prescribed questions, but Keisha generally stared straight ahead and refused to answer. She clearly hated the whole exercise and eventually refused to come.

I didn't blame her. She was missing so much information about the world that focusing on “strategies” seemed beside the point. When we read a book about the Golden Gate Bridge, for example—to practice “summarizing,” or whatever the strategy of the week was—I discovered she'd never heard of California, let alone San Francisco or Marin County, terms used in the book.

One day Keisha said softly that she really should be in class, where she might be learning something (instead of in a one-to-one tutorial). I had to agree.

That's [Natalie Wexler](#), a wonderful education writer, describing her own recent tutoring experience. Hers is a common story, not an outlier.

Natalie's example was with one student. Frequently there are two to four kids in a tutorial. Some are not paying attention, eyes drawn to the windows, or covertly scanning phones in their laps. Some are

confused, brows furrowed. Some are irritable, lips pursed, sighing theatrically at the slightest challenge, rolling their eyes at each task change. Often the tutor is talking too much, over-explaining.

“

It's quite hard to create a productive tutorial... the work of well-intentioned adults too often leads to worthless lessons.

”

Often both parties are bored with the curriculum. Maybe the Zoom session just timed out. Maybe Kid 1 was tight with Kid 2 in September, but they had a big fight, and now they hate each other. Maybe Kid 3 was “sort of OK” with tutoring but became resentful when the tutor called her mom and mentioned missing assignments.

It's weird in there.

If you believe tutoring is simple, then the path to scale is easy: Get some cash, use it to pay smart and kind adults to

sit across from kids and teach, create some rules, and get to work.

If you think “it's weird in there”—that it's quite hard to create a productive tutorial, that the work of well-intentioned adults too often leads to worthless lessons—you'll grasp that it would be easy to waste a ton of money in this arena.

But as we'll explain in the next section, it's not just that individual tutorials can be quite challenging.

II. Cells and schools constantly create *new* problems

Cells change. When you make a vaccine, you have to account for that. [Derek Lowe](#) writes about how this makes vaccine-making hard:

That ribosomal complex is surrounded by other proteins whose job is to constantly look for trouble and try to correct it...

How about transcription? You might have a model of a few ovals and circles (the RNA polymerase complex) smoothly ratcheting along a strand of DNA. But in reality, many of those complexes take off in the wrong direction and stall. And they will ratchet backwards on purpose if a base polymerization error is detected, although that process sometimes stalls out, too. This “backtracking” has a whole suite of correction mechanisms on top of it.

Schools change, too. New schedules. New leadership. New priorities. New internal politics.

All of these changes can send a high-dosage tutoring program spinning. These challenges need to be straightened out, constantly, or students and tutors get frustrated. School changes pile up on top of each individual tutorial change.

“The assumption is that the school will more or less stay the same, so you can stick to the recipe. That assumption is often wrong.”

Most education programs, including tutoring, try to create a recipe, or a list of best practices. To scale, they declare that the recipe “works.” The assumption is that the school will more or less stay the same, so you can stick to the recipe. That assumption is often wrong.

But that’s not the approach of the best tutoring programs. Successful leaders know that schools will change and things will jam up. It’s inevitable. So they hire unusual managers. Not people who follow a recipe, but people who solve problems. People who hope to preserve the recipe but expect to adapt it.

These unusual managers obsessively look for problems caused by school changes, fiercely try to fix them, and humbly realize that often their first and second and third “fix attempts” might not work. They persist until they get the right result. Here we see the link connecting the organizational culture of the service provider to what the students experience in the tutoring session.

Typical managers in education aren’t like this. They don’t obsessively seek to find problems. In fact, they fear finding problems. From above, the “big shots” insist that they have the right model, blinded with confirmation bias, seeking only comforting data. Managers learn to hide reality. When problems eventually spill into the open, it triggers a fight-or-flight mechanism. They either fight the people

describing the problems (often tutors), or in taking flight, they create a new policy that nominally solves the problem, then they declare victory and move on to the next thing.

Leadership is the desire, aggression, humility, and persistence to identify and actually fix problems—which, in schools, never stop coming.

To the best of our knowledge, there is no clear research on the effect of “tutor program leaders” on tutoring outcomes. It is interesting, however, to look at charter schools. Most of the outlier successes, like KIPP, which have replicated [successfully](#), do not rely solely on their “program.” They focus instead on the recruiting and training of *unusual* school leaders, their “Fisher Fellows.” Specifically, they seek leaders who will succeed *downstream*—when their schools and their operating contexts change in unpredictable ways. (For example, when a pandemic requires a rapid shift to online instruction!)

And while KIPP-with-carefully-chosen-leaders has scaled up well, KIPP-just-the-program has not. We hypothesize that the same is true for start-up tutoring programs.

Just like with cells. Winning vaccines anticipate cell changes.

III. A respectful disagreement about research

Bob Slavin, a wonderful researcher, has written some hard truths about Covid-19 learning loss. He correctly dismisses policy interventions like extending the school year, typical after school programs, and summer school. Those won't work. Bob believes tutoring, however, [will work](#):

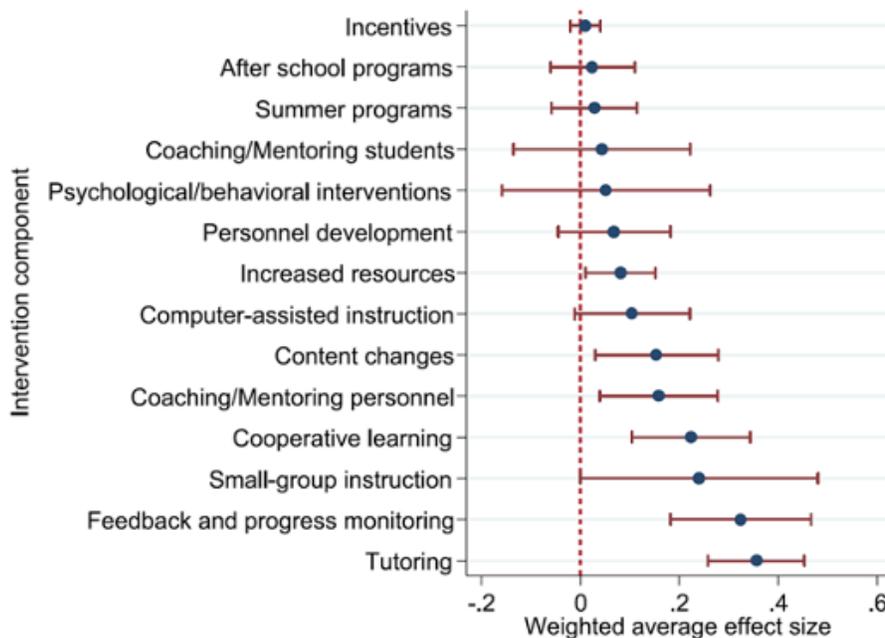
By far the most effective approach for students struggling in reading or mathematics is tutoring (see blogs [here](#), [here](#), and [here](#)). Outcomes for one-to-one or one-to-small group tutoring average +0.20 to +0.30 [standard deviations] in both reading and mathematics, and there are several particular programs that routinely report outcomes of +0.40 or more. Using teaching assistants with college degrees as tutors can make tutoring very cost-effective, especially in small-group programs.

Effect sizes are a wonky way to describe impact. Our friend Matt Kraft, for example, writes that a 0.2 standard deviation effect in education is large.

Bob has also written [a Marshall Plan](#) for scaling up tutoring, a document we admire.

Philip Oreopoulos and his colleagues agree with Bob. They recently published a meta-analysis, with very large positive effects—0.37 standard deviations—across scores of tutoring programs, some “high dosage” and some not.

And [Dietrichson et al.](#) offer the following chart in favor of tutoring, from 2017:



Clearly, there is reason for scholars to be optimistic about tutoring. And we are pleased that the programs with which we've been affiliated are included in that positive research. We're glad that Experience Corps, Reading Partners, and the i3 study of Reading Recovery had fantastic, quality randomized control trials, and that impressive evidence showed gains for students.

So why worry?

We worry because we believe that many more tutoring programs fail than is commonly believed. There are two key drivers of that belief: publication bias and scale-up problems.

Publication bias

This is a broad problem that affects more than tutoring. Many failed education programs never stick around long enough to get measured. For example, when we scaled high-dosage tutoring from Boston to Houston, Texas, it worked. But missing from the story is that the Austin, Texas public school system tried to create its own version of HDT around that time. It quickly failed and disappeared without a trace. That's not included in the research.

When Saga brought HDT from Houston to Chicago, their program succeeded. But at the same time, another Chicago school network launched its own HDT program. That one died after nine months. That's also not included in the research.

“ [Publication bias] is a broad problem that affects more than tutoring. Many failed education programs.... [are] not included in the research. ”

When Mike did HDT in our Boston charter school, we deployed literally those same tutors into nearby district schools. The charter students received large gains; the district students had no gains. Indeed, every time we've been part of a successful tutoring program, one which “enters the tutoring scholarly literature,” we've seen a similar program fail, yet disappear too fast to ever be captured.

Imagine a group of people who try an experimental drug, have bad outcomes, but nobody notices the bad ones, only the good ones, so the drug overall seems successful. (You don't have to imagine too hard: hydroxychloroquine for Covid-19).

There are presumably hundreds of such tutoring failures. We believe the positive-research story omits these.

Scale-up problems

With vaccines, the 1 millionth dose is identical to the 2 millionth dose. But that can't happen with education programs delivered by human beings.

In particular, education programs that work small often don't work when they get big. A famous example: Fifty-eight kids who benefitted from the oft-cited Perry Preschool Project in 1962 didn't translate so well to 18,000 children in [Tennessee](#) or to a similar program in Quebec. For a more recent

example: Last month our friend Ben Feit published a [study](#) finding that Texas charter schools did not scale up well.

Tutoring—not high-dosage tutoring, but what we might call “regular tutoring”—already failed in a big national scale up: the George Bush/Ted Kennedy No Child Left Behind version. Again, it’s hard to [get the details right](#).

Every education intervention, including tutoring, requires a newly formulated, carefully calibrated program. Its details ought to depend on which students it serves, whether it’s optional or mandatory, which tutors it uses, what time of day it’s offered, whether it’s online or in person, what the tutor-student ratio is, what curriculum is used, what the leadership is like, the overall school culture, and more. There is no de-situated, de-contextualized “thing at rest,” no vaccine. Many have [written](#) about the need to consider if a program is “hard to scale.”

“The evidence on tutoring is, we believe, choppier than we would have wished.”

The evidence on tutoring is, we believe, choppier than we would have wished.

So what should we do?

IV. How to move forward: Vaccine development and productive “failed trials” as a model for scaling up high-dosage tutoring

Vaccine development goes like this.

- A.** Each new situation, or disease, requires a newly formulated vaccine. You don’t start 100 percent from scratch. You get to take your best ideas from previous vaccines. But the new vaccine candidate is measured anew, as it fights in a new context.
- B.** Even proven teams—chock full of MIT and Stanford Ph.D.s and their equivalents—fail a lot. Failure happens because “it’s weird in there.” There are so many interconnected moving parts. If there are ten challenges but you only solve nine, the vaccine fails.
- C.** Many vaccines are attempted, measured, and abandoned. The culture of the pharmaceutical industry is at its best when it embraces this frequency of failure.

As the [New York Times](#) wrote, “Nine drug companies issued a joint pledge that they would ‘stand with science’ and not put forward a vaccine until it had been thoroughly vetted for safety and efficacy.” Even though people are dying, the pharmaceutical industry and Western governments will *not* just throw money at an unproven vaccine and risk public health and trust.

To the talented teams that work hard at a vaccine and fail, we say, “Thanks for trying.” To the patients who’d really benefit from a vaccine, we don’t give them a low-quality vaccine. We just say, “Unfortunately, we don’t have a winner yet.”

But things are different in Russia and China. They approve vaccines based on what worked with *previous vaccines*, but have *not* been proven in the new context of this particular virus. China began immunizing people in Jiaying back in October. Russia rolled out their vaccine without a randomized controlled trial. This *can* work, yeah, but Vladimir Putin [won’t take it](#).

That’s typically our education scale-up strategy. “X worked before. Y, with the same program but new leaders and a new context, will work, too.” We act like Russia and China, not Western countries.

Our recommendation, therefore, is to scale up tutoring like we do vaccine trials.

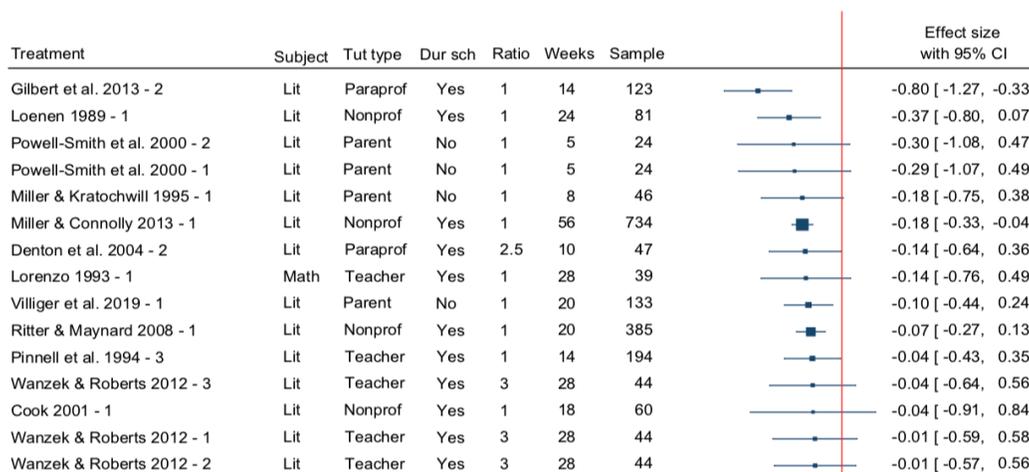
How do you get great high-dosage tutoring to students who need it? Many well-intentioned HDT programs should be attempted, measured, and abandoned. The successful ones should be grown appropriately. That’s how vaccines work.

There are two primary steps for replicating this with high-dosage tutoring. First, fund “HDT Candidates,” and fund “HDT Programs”

With vaccines, teams get money to *try* to create vaccine *candidates*. Just because once before you once made a winning vaccine for Ebola doesn’t mean you can just whip up a Covid-19 vaccine and produce it.

That’s what we should do with tutoring: Make it easy for previously proven teams to create customized HDT candidates in specific settings. If it’s three hours a week of retirees tutoring struggling second graders in phonics, virtually, with some specific curriculum, in the Atlanta Public Schools, great. Try that. Phase 1 should always be a low bar. Something like 200 students randomly assigned to control and treatment groups.

Let most HDT candidates die after the evaluation, just like vaccines. Lots of very well-intentioned programs fail. [From Andre Nickow, Philip Oreopoulos, and Vincent Quan \(2020\):](#)



These programs had negative effects! And those are programs that stuck around long enough to be measured.

Imagine all the programs so obviously bad that they disappear without ever being measured. Embrace that reality. Let vaccine candidates and HDT candidates be created, planning all along that most will fail, so you have to measure their effects carefully and pull the plug on them when they do.

The second primary step in finding successful programs and growing them is to offer a choice to HDT programs that do work in small settings. One, they can choose to “conditionally scale” to, for example, 1,000 or 5,000 students. But the idea is that we know that their programs might dilute, that they still need to be measured with an understanding that there is a good chance the “scaled version” fails even as the “small version” succeeded. Or two, those successful small programs should be able to choose to stay small. Don’t force growth.

“HDT trials” would cut sharply against the dominant education culture, which is usually to prescribe the inputs. “Here are forty-seven pages that describe what you must do,” a typical policy might mandate. “We know what to do, now you do it.”

But that's wrong. We do not know what to do. Adaptations will be required. There is no "thing at rest."

Instead of prescribing the inputs, describe the results you're looking for. Define the outcomes. Then free educators to embrace that "it's weird in there." A non-negotiable condition is finding *unusually proactive managers* to deal with the reality that schools, like cells, change.

Also give separate funding for outside researchers to measure the results, allowing tutoring providers

“

Instead of prescribing the inputs, describe the results you're looking for.

”

to pick from pre-qualified research teams. And consider the market side, too, [proposed by Paul DiPerna](#).

The real risk is opportunity cost! Doing this the "normal way" will put HDT in the same K-12 graveyard as every other education intervention that had its fifteen minutes of fame, then didn't do well when

carefully measured at scale, and now is discredited.

Free education leaders to embrace failure, like scientists do, instead of insisting, like educators do, that they've succeeded and arranging the evidence to look that way. Yes, many educators want the Putin way, or the China Way. "Here is our proven vaccine," they claim. But the truth is that the Russian and Chinese Covid-19 vaccines have not been through all three standard RCT phases. Will you take it? Give it to *your* kid?

Coda, February 21, 2021

Since we first published this essay in blog form, there have been a few notable developments.

1. On vaccines: Big picture, Covid-19 vaccines are a success story. By sticking to the scientific method, we're starting to bring the pandemic under control.

Importantly for our essay, the three biggest, most established vaccine makers [failed in their quest](#) to create vaccines: Merck, GlaxoSmithKline, and Sanofi. They took their “proven” (in other contexts) ideas, tried them on Covid-19, and failed.

2. The inestimable Bob Slavin offered his take on our original blogs. We reprint his comments here with permission:

In principle I agree with your suggestion that Phase 2 should be based on Phase 1 findings or experience. But as a practical matter, that is impossible.

Phase 1, which will probably have to be funded by foundations rather than federal money (because federal money would take too long), would last only eight months, February to August at best, and it would be building capacity and scale up to the last minute.

It is true that we will be learning quite a lot even in the spring, and these learnings should be reflected over time in improved policies and procedures. But the federal legislation has to be based on the best assumptions available roughly now, or very soon.

Everyone is eager, with good reason, to get whatever program is selected off the ground as soon as possible, by September 2021, if even that is possible, and there is no possibility of a pause to consider learnings from the Phase 1 experiences before launching Phase 2.

My conception of this is that with the Phase 1 pilots and the 2021–2022 evaluations, the overall program will continually improve its reach and impact. The evidence base for tutoring is already good enough to predict positive outcomes, at least for elementary reading and math, so there is good reason to go ahead with what we already have, with a process in place to get a lot smarter over time, or so I believe.

In comparison to, say, SES (Supplemental Education Services—the tutoring component of No Child Left Behind, where parents could in theory choose tutoring providers besides their home district), we would be starting miles ahead, with already proven programs and a process to create and evaluate additional programs and learn from early implementations that neither SES nor any other large federal program ever had.

Federal education programs usually start with untested ideas and end with an autopsy, learning little or nothing along the way.

We agree with Bob on the big picture. Details Matter. Money alone won't help here. It would be crazy for local officials to simply take federal or philanthropic funds, throw something reasonable together, and launch it, rather than use tested programs. We're just more skeptical about the tested programs with new people in new contexts, and whether they'll replicate.

We also agree that tutoring research shows much more powerful effects than most other education ideas that have been scaled up, such as various forms of teacher training, curriculum, extended school day, teacher evaluation, and more.

And we concede that Bob may be right that our proposal might be laudable but unrealistic. Perhaps we're letting the great be the enemy of the good.

“We're just more skeptical about the tested programs with new people in new contexts, and whether they'll replicate.”

3. We had conversations with public officials in USA and the Netherlands. American leaders expect two big new investments of federal stimulus: the “Trump” one that states have already received, and a forthcoming “Biden” one. Some of that, they say, they'll spend on tutoring to combat Covid-19 learning loss in September 2021. In addition, Americorps could see a huge increase, with some of those dollars meant for full-time tutors. Meanwhile, across the pond, Dutch officials have just announced similar investments.

Money will be there.

But what about *capacity*? School system leaders are exhausted—“fried,” to use a common word. Moreover, they have little capacity to think about September 2021! They have so much on their immediate plates: openings and closings, acute issues around cultural competency, a turbulent budget season ahead this spring, and “old” priorities (*way* back from February 2020).

The way forward? We say they should launch all new efforts with rapid measurement, without too much red tape; pull the plug without blame on some tutoring efforts, instead of pretending they'll be “fixed” with a few tweaks; and double down on those efforts that are succeeding—all while supporting and communicating about what works, and being honest about what doesn't, and doing so with maximum transparency and humility, with the least amount of spin, and without burying negative results.

Just as our success with Covid-19 vaccines is rooted in the failures of many “proven providers,” so too can high-dosage tutoring help many children—if and only if we're willing to be humble and avoid focusing on inputs, and react instead to outputs—to how students X respond to particular tutoring Y in particular contexts.

About the authors

Mike Goldstein is the founder of Match Education in Boston: a college prep charter school for low-income kids; an embedded Graduate School of Education; and a program to share best practices. He is the former Chief Academic Officer at Bridge International Academies, an organization that has launched or operates over 900 elementary schools in Kenya, Uganda, Liberia, India, China, and Nigeria. Mike has been an advisor to various organizations, including Schoolhouse (micro schools), Croft Schools, Reconstruction, Avela (school choice), Future Ed at Georgetown University, National Council on Teacher Quality, Harvard Education Press, Transcend, and more.

Bowen Paulle taught for six years in high poverty secondary schools of the South Bronx (New York) and the Bijlmer (Amsterdam). Since the 2013 publication of the comparative study based on his time in these two fields (*Toxic Schools: High-Poverty Education in New York and Amsterdam*, University of Chicago Press), Paulle has focused on potentially scalable interventions meant to bring about breakthrough outcomes for disadvantaged students. Paulle is committed to a utilization-focused evaluation paradigm combining ethnographic detail and research-based advising with the rigor of randomized controlled trials.

Suggested citation for this report:

Michael Goldstein and Bowen Paulle. *The narrow path to do it right: Lessons from vaccine making for high-dosage tutoring*. Washington D.C.: Thomas B. Fordham Institute (March 2021).

<https://fordhaminstitute.org/national/research/narrow-path-do-it-right-lessons-vaccine-making-high-dosage-tutoring>