Field officer Imisi Olu-Joseph of Bridge International Academies conducting observation at Oba-Ovonramwen Nursery and Primary School in Benin City, Edo State.
Beyond Moneyball: Data-Driven Education Boosted by Observation and Judgment

In Nigeria, smiling at small victories and pushing to do better

By MIKE GOLDSTEIN
WHEN I WORKED for Bridge International Academies, the largest network of elementary schools in the developing world, gee, did we have a lot of data. We had testing data from five different countries, each with its own national curriculum and practices. (Kenya, for instance, administers standardized tests not only in math and English but also in science, social studies, and Kiswahili.) We had data on teacher observations, parent satisfaction, peer tutoring, and even parent-teacher conferences. It was a huge amount of material, even for a data nerd like me.

As chief academic officer, part of my charge was to sort through and make sense of all this quantitative information, with the end goal of improving instruction and student outcomes. Bridge operates both low-cost private schools, somewhat akin to American inner-city Catholic schools, and public-private partnership schools, similar to Obama-era turnaround schools. More than 800,000 students are enrolled in 2,026 schools in five countries. When I talk to friends who run charter management organizations or school districts in the United States, the question they always ask me about Bridge is: “How did you use the data?”

I start my reply with a word of advice: When taking on a new education venture that you intend to evaluate, reach out to top-notch economists who can measure your results through a randomized controlled trial. That will give you solid information on whether you’re actually helping kids make significant gains, and it will help you avoid the very human tendency to “believe what you want to believe” when you look at achievement data. Much of it is related to how the scripted lessons, or teacher guides, are used in class: In grade 6 math, the first example given—”I’ll pay the price for two. I don’t want a second rider seated behind me.”

They negotiate, settle on a fare, and take off. “Sometimes louts try to stop the bike and collect ‘tolls,’” Imisi tells me. The taxi arrives at a smaller bus station, Igando. There Imisi will hire a second bike for the final stage of the journey, to the school at Dare Olayiwola.

The school manager is “attending to parents,” when Imisi arrives. “I won’t chat him up,” she tells me, “just say good morning, smile, and pass. I want to stay in the shadows. I’m here to observe for several hours.”

What does Imisi observe today?

Much of it is related to how the scripted lessons, or teacher guides, are used in class: In grade 6 math, the first example given by the teacher was unclear. The grade 5 English class teacher tried to cover way too much material, and the kids were confused. The grade 1 teacher got tripped up in the science guided practice.

Also: One class was short on math textbooks, so the teacher had to write the examples on the board; more books were delivered to her shortly thereafter. A field team observer rated a recent parent-teacher night a 7 out of 10, when an earlier version had earned a 4 out of 10. The school staff thinks it went better this time because teachers led with a personal anecdote about the child rather than launching straight into an explanation of grades.

It’s mostly “little stuff,” but it adds up.

Bridge field officers don’t inspect teachers; they collegially critique the work of the senior officials—the directors of training, instructional design, technology, and operations, who are not allowed to “blame implementation.”
There is also an experiment underway at the school, using an MIT-validated technique in which students from various grades are regrouped by skill level. Imisi is closely watching how students seem to feel about being with older or younger classmates. Does anyone look intimidated? Or ashamed?

Imisi watches a class of 20 students, in which every child except one can do the lesson. The teacher gets frustrated and raises her voice. Here, Imisi steps out of the shadows and models a patient approach. “I’d seen that girl succeed in another class,” Imisi tells me. “The teacher just needed scaffolding.

The View from the Ground

Every CEO, every general, every school superintendent needs to know: What is really happening on the front lines?

When a top official visits a classroom or a school, people notice the Big Cheese and change their behavior. When an official asks for information, the answers are often what the respondent thinks the official wants to hear, rather than an account of what is really going on. This phenomenon is not exclusive to schools. For example, it’s the dominant theme of David Halberstam’s *The Best and Brightest*, about the Vietnam War, a time when JFK and LBJ were unable to hear the true story.

In the Western world, the K–12 sector usually tackles this puzzle in two ways.

In the UK, there are “inspectors.” They arrive with long checklists and good intentions, and fan out to classrooms. "Does the teacher show high expectations?" Inspector raises head from notebook to watch Ms. Smith. Ms. Smith calls on a student. The child doesn't know the answer. Ms. Smith quickly moves on. Inspector etches a red mark to indicate low expectations. The
report eventually goes to the school leader, who often scolds the teachers. (Sometimes, of course, the inspector misinterprets the scene. “If I persist with questioning that particular child in front of his peers,” Ms. Smith might have said, “experience tells me he’s going to blow up in anger. If the inspector had waited, he’d see I helped this particular boy after school. That’s our agreement.”)

The other prominent K–12 effort to grasp what is really happening in a classroom is teacher evaluation, common in the United States. This approach nudges principals out of their office chairs to show up in class and watch, then write up feedback. Despite the best of intentions, this policy effort hasn’t gone as planned. Principals don’t want conflict. The observations take time. And principals are not that good at the business at hand. The Gates Foundation Measures of Effective Teaching Project found that students, using a simple survey developed by Harvard economist Ron Ferguson, were two to three times more accurate at rating teachers than principals were. Principal evaluations were dangerously close to having no correlation with student learning gains.

Sometimes the principals just validate their own style of teaching. Other times they reluctantly fill out scorecards, which tend to focus on teacher actions, not student learning. This is a common failing in education, rating the inputs instead of the outputs. It would be as if you rated a baseball batter on swing aesthetic—does it look pretty?—rather than on-base percentage or runs batted in. It’s misleading and irrelevant. I recall a teacher at a Boston charter school who was known for a terrible “aesthetic.” He never seemed to be trying, defying all the observation rubrics. He just sat there while his students read books. Yet his kids made large gains on the English exams.

Imisi isn’t out to verify any theories. Her approach to classroom observation differs from the traditional kind in four main ways.

First, instead of sending her out to validate a hypothesis that headquarters hopes is true, Bridge sends her there to reject a hypothesis. She goes in assuming the lessons are failing, that often students are daydreaming, that the pacing is off target. In the United States, many classroom observers in public schools describe feeling pressure to say things are going well. Imisi and her colleagues are nudged in the other direction: there is a ton to be fixed; please go find it.

Second, the Bridge brass doesn’t just want Imisi to fill out a rubric—they also want her overall judgment, her big-picture take on things. On a scale of 1 to 10, to what extent is a lesson or a pedagogical approach or a tech tool succeeding with students and teachers? By contrast, officials in the United States seldom ask the big-picture questions; the observations are all forced into preexisting categories.

It’s fine, even desirable, for Imisi to focus on the little ungla- morous things. For example, at two minutes and 54 seconds into the lesson, the teacher’s instruction to the kids to break into small groups was confusing, so they just stared at each other. A set of three math problems was meant to take five minutes, but even the speediest kids needed 12. That messed up the timing of the whole lesson, and as a result, the teacher didn’t get to the quiz.

(Timing is everything. Back in 1992, one of my first jobs was working as a gofer for a Broadway theater producer. I recall a rehearsal—I think the show was Guys and Dolls—in which Jerry Zaks was directing Nathan Lane. They were fixing a line that had flopped in previews. “Pause after you say it for five seconds, not two,” Zaks told Lane. That was it. The next night, the laughter started to build around the four-second mark, and it killed. In the United States we like to argue about the lofty questions of rigor, when often what makes or breaks a lesson is pacing.)

Third, Imisi’s “target” is different. School districts commonly use inspections as a way of critiquing teachers. Imisi is not inspecting teachers; she is collegially critiquing the senior officials—the directors of training, instructional design, technology, and operations. They are not allowed to “blame implementation,” a common phrase in ed reform, which essentially says, “I think I created a magical tech tool or lessons or coaching, but geez, our teachers just mess it up.” That doesn’t fly. The tools, training, and lessons are designed to be used by mere-mortal, typical teachers. If they aren’t using these resources well, or they’re rejecting them, it’s “on you,” the senior official at Bridge. Do better. As Yoda said, “There is no try.”

Imisi serves up helpings of forced humility to senior officials on the team. She is unsparing. She might rate a lesson a 3 out of 10. That stings a curriculum director who worked hard on it. But maybe, over time, with a fail-fast mentality, the director will manage to improve the lessons and earn a 4 and then a 5, perhaps eventually reaching an exalted 6 out of 10. Imisi acknowledges that the kids are “getting” it, that the lessons are more or less working, but they are far from masterpieces.

This stands in sharp contrast to some curriculum efforts I’ve observed or been part of in the States. For example, at one midwestern charter school, I saw a class I’d describe as a 3 out of 10. The school had just adopted a new Common Core curriculum, so the low score was understandable. When I shared this with a friend back at the curriculum company, she blamed

---

Based on Imisi’s reporting, someone at headquarters might ask her:

**Tomorrow, can you zoom in on this nuance, shoot video of that detail, ask teachers their view on this possible new direction?**

---

Feature • Bridge International • Goldstein
the teachers. (“Yes, we often see low expectations from teachers; it’s sad,” was the response.)

Fourth, the relationship between headquarters and the Bridge field officers is dynamic; information flows both ways. Based on Imisi’s reporting, someone at headquarters might ask her: Tomorrow, can you zoom in on this nuance, shoot video of that detail, ask teachers their view on this possible new direction?

Reverse Moneyball

Data nerds love Moneyball. In the 2011 movie, Brad Pitt plays Billy Beane, general manager of the Oakland A’s. It is fall 2001, the team has just lost to the Yankees in the playoffs, and Beane has to rebuild his roster as he faces the imminent loss of three superstars. His payroll budget is slim compared to the financial-powerhouse Yankees and Red Sox. How to win? Beane, a former player himself, begins to reject the wisdom of his veteran scouts and even his own knowhow.

Cognitive bias, he learns, leads to misperception. Chad Bradford looks like a bad pitcher because he throws underhanded. But he’s actually great at getting hitters out. He’s worth a lot in terms of generating wins, but so far, nobody realizes it, so he isn’t paid much. Beane picks him up at a bargain rate.

Beane meets a brilliant Yale economics grad, played by a sweaty Jonah Hill in a polyester navy blazer, and hires him to lead a data-analytics revolution at the baseball franchise. Cold numbers replace hot human opinions, and newer, more consequential numbers replace outdated ones.

Old-timers resist Beane’s new approach. The head scout quits. The manager accuses Beane of sabotage. But data eventually triumphs. (The A’s lost in the 2002 division series, but they won 20 consecutive games in the regular season, setting an American League record.)

Data analytics soon swept through pro sports: Theo Epstein and the Boston Red Sox, Stephen Curry and the Golden State Warriors, Bryson DeChambeau and his data-driven path to professional golf excellence. Data haven’t entirely conquered sports, but they have secured a place alongside human judgment and experience—sometimes weighted a little more and sometimes less.

In the realm of American ed reform, George W. Bush and Ted Kennedy arguably ushered in the Moneyball era in 2001 with No Child Left Behind. Some seemingly bad schools were actually good, if you accounted for student starting points. The teachers helped their students learn more than similar kids were learning in other schools. Some apparently good schools were not, if you took a careful look at the performance of “subgroups”—poor kids, minority kids, special ed kids.

Ed-reform analytics caught on quickly. Old-timers resisted. But in this battle, data lost. “Data-driven instruction,” data-validated Common Core curriculum, data-driven leadership, school turnarounds, and teacher prep: by and large, they have not worked. Yes, there are worthy exceptions. Some charter schools, perhaps the D.C. public schools for a while, have achieved data-driven success. These outliers were supposed to be the Oakland A’s, in the vanguard. If you build a better mousetrap, it’s supposed to be copied.

That hasn’t happened in any meaningful way in America’s public schools. Academically, poor kids are more or less where they were 20 years ago.

Some critics think “data-driven reform” isn’t the A’s or the Red Sox; it’s the 2017 Houston Astros. It’s cheating. So of course things haven’t improved.

Some reformers think they’ve been defeated by sheer political power, that no matter how convincingly the data speak for a particular success, the powers that be are always moving the goal posts, and the winning ideas are not allowed to spread. And many observers think that a lot of the people who run school systems value votes over student achievement gains.

I have a different take, or maybe an additional take: In the United States, analytics work in ed reform just hasn’t been that good. The numbers crunching has added up to . . . meh. It’s missing something.
Bridge International Academies is *Moneyball* in reverse. In the K–12 world, data are already omnipresent, but they are mostly misunderstood and misinterpreted, mangled and misused. Bridge’s unsung field officers provide human judgment. Putting that alongside big data is the secret sauce. That, I believe, is the *something* that has been missing.

We need a new breed of human judgment in schools—not the old intuitive kind of judgment, as in wise elders proclaiming, “I am experienced, and here is what I believe and feel.” We need trusted neutral observers, wise to be sure, but simply and consistently narrating what is going on rather than instilling their pet beliefs, and willing to have their own narratives critiqued and evaluated. Such observers are not expected to solve the problems themselves—that might motivate them to see what they want to see. Instead, these observers can constantly fine-tune the information provided by big data, and the data can, in turn, guide further observation.

But I’m getting ahead of myself. Let me describe Bridge’s two all-beef patties, lettuce, cheese, pickles, and onion on a sesame-seed bun. Then we’ll get back to the special sauce.

**Bridge**

I worked for Bridge International Academies from 2013 to 2016. I remain involved as an adviser. Notwithstanding the many imperfections of Bridge, including and especially my own errors, I believe overall the organization is a good thing, even a remarkable thing. My wife and I are close with four Bridge alums—Natasha, Grace, Josephine, and Geof, all of whom are now scholarship students at American colleges. In fact, I am dictating these words as I drive to Bowdoin College to pick up Geof to hustle him to Boston’s Logan Airport tomorrow so he can return to Nairobi for the first time in two years. Seeing those four kids thrive and flourish has been a great joy; they’ve seized the opportunities they’ve been given. It makes me wish so hard there were a way to unleash all that latent potential in all of Bridge’s 800,000 children, and the hundreds of millions more in the developing world. Failure to do so is an epic waste.

Bridge is akin to a charter management organization combined with a turnaround organization. In its low-cost private schools, parents pay about $100 a year for tuition. (Schools like these serve hundreds of millions of children around the world. See “Private Schools for the Poor,” *features*, Fall 2005.) It also works with “turnaround” public schools, in which the government contracts with Bridge’s parent company, NewGlobe, for instructional materials, training, and expertise. These schools are not called Bridge academies, but it’s the same academic team participating.

Operating in Kenya, India, Nigeria, Liberia, and Uganda, Bridge is larger than the 10 biggest American charter management organizations combined. It’s organized as a for-profit company with a public-good mission and is backed by a number of “double bottom line” investors, including Bill Gates and Mark Zuckerberg.

Bridge presents as an ed-reform inkblot test.

If ed reform makes you think attrition, expulsion, colonization, teaching to the test, privatization, and treating teachers badly, you probably won’t like Bridge.

If you tend to like charters, choice, and “parent power” of various types in the American context, you might lean into Bridge.

My purpose here isn’t to persuade you that Bridge is good or bad. My purpose is to tell data lovers that we dramatically underinvest in the companion field of observation and, as a result, don’t have the big helpful effect on schools that we might.

**The Developing World Context**

If all you know is American education, here’s some perspective.

The United Nations take on education in the developing world was, until recently, that not enough kids were in school.

More recently, the consensus is shifting toward a different problem: learning outcomes have always been poor, and they seem to be getting worse.

Beyond the sticky problem of getting children into schools lies the challenge of getting teachers to show up and persuading them to stop using corporal punishment. Lant Pritchett, a development economist, has written, in reference to Indian schools, that even when teachers do show up, they might not bother to do their jobs.

“Less than half of teachers are both present and engaged in teaching on any given school day,” Pritchett wrote, “a pattern of teacher behavior that has persisted despite being repeatedly documented.” What’s more, Pritchett noted, a survey of Indian households “found that about 1 out of 5 children reported being ‘beaten or pinched’ in school—just in the previous month.” The study also found “that a child from a poor household was twice as likely to be beaten in a government school as was a child from a rich household.”

Entrepreneurs Jay Kimmelman and Shannon May, who opened the first Bridge academy in Kenya in 2009, jumped on those challenges. Their low-cost private schools had high teacher attendance compared to competing nearby schools. They fired teachers who used corporal punishment (even though some parents *liked* it).

**We need trusted neutral observers in schools, simply and consistently narrating what is going on rather than instilling their pet beliefs, and willing to have their own narratives critiqued and evaluated.**
Pritchett also wrote of an Indian study in which observers visited classrooms to look for “any of six ‘child-friendly’ pedagogical practices,” such as “students ask the teacher questions” or “teacher smiles/laughs/jokes with students.”

“In observing 1,700 classrooms around the country the researchers found no child-friendly practices at all in almost 40 percent of schools—not a smile, not a question, nothing that could be construed as child-friendly engagement,” Pritchett reported.

This one is more complicated.

For years, USAID and agencies in other countries spent huge sums to train teachers in the developing world. Yet careful empirical evaluations rarely found that training efforts alone would raise student achievement.

It was Benjamin Piper who cracked the code for changing teacher behavior. Piper is a longtime doer and scholar at the Research Triangle Institute, or RTI International. His USAID-sponsored projects have sparked big changes in elementary education in certain developing nations.

Piper realized that training alone couldn’t change teachers, because they themselves had attended schools where teachers relied solely on lecturing and rote call-and-response pedagogy. So they had developed a strong inclination to teach that way, too, notwithstanding any professional development from Western do-gooders.

Only scripted lessons—which blocked teachers from their default practice of lecturing even small children for very long durations—seemed to change the classroom dynamic. The scripted lesson or teacher guide is a coercive tool used for a liberal end, essentially forcing teachers to say something like: “Now I am going to stop talking, and you students are going to . . .” read, or write, or talk with one another.

Once Piper had these teacher guides in place, he could layer in highly focused training on how to succeed with this particular style of teaching. Resources mattered, too: students needed real books to hold and read (not an easy thing to provide in many corners of the world).

Those three things—scripted instruction, focused training, and essential resources—added up to Piper’s Primary Math and Reading (PRIMR) program, and later one called Tusome (“Let’s read,” in Kiswali) in Kenya and Tanzania. The student learning gains arising from these programs are impressive (see Figure 1).

The teacher guides are an understandably touchy point. An Atlantic article about Bridge (and not about Piper) is headlined “Is It Ever Okay to Make Teachers Read Scripted Lessons?”

Author Terrance F. Ross wrote that the uniformity of the lessons “all but guarantees consistent results,” but:

... by its nature, this approach stymies individuality and spontaneity. Dynamic educators who are adept at innovating on the fly and creating unique classroom experiences don’t necessarily exist in the Bridge system. They are eschewed in favor of teachers who can follow instructions well. Bridge’s argument seems to stem from a utilitarian philosophy: Based on Kenya’s dismal public school statistics, it’s better to give all children a basic, reliable education than hope for talented teachers to come along.

There’s great merit to the notion of teacher freedom, but the teaching in non-Bridge, typical Kenyan schools is not based on...
What Worked? What Didn’t? *(Table 1)*

Bridge tries lots of different ideas and initiatives, many of which are backed up by research in other school settings. The question is: will they work at Bridge? The first test is whether Imisi Olu-Joseph’s team observes the idea working well. The second is whether the data show learning gains, as measured by a randomized controlled trial if possible.

<table>
<thead>
<tr>
<th>INITIATIVE</th>
<th>RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Interleaving” math lessons in Nigeria: mixing in topics from weeks and months earlier, instead of hitting each topic once during the school year</td>
<td>Success at scale</td>
</tr>
<tr>
<td>Assigning math problem sets of varying levels to kids in the same classroom in Kenya</td>
<td>Failure</td>
</tr>
<tr>
<td>Kindles for pleasure reading in Liberia</td>
<td>Success, <em>but . . .</em> much of the support for this undertaking was provided by another NGO, and when the pilot expired, the project ended. This was personally disappointing to me, because most children in the developing world have no pleasure-reading books at home or at school and no access to local libraries.</td>
</tr>
<tr>
<td>Peer tutoring (6th graders tutoring 1st graders)</td>
<td>Success with 20-school pilot, <em>but a failure at scale when measured with randomized controlled trial</em>. The trial did show a 0.06 standard-deviation gain in math, but the effect was null in English. Still, many parents liked the project.</td>
</tr>
<tr>
<td>Weeklong academic enrichment camp in Kenya, leading up to high-stakes national exam that determines high-school admissions for the whole country</td>
<td>Success in pilot version with some truly exceptional teachers, <em>but a failure when scaled to a larger size</em>. This is a familiar phenomenon. Many education interventions, such as high-dosage tutoring or teacher coaching, succeed as artisanal efforts powered by a small group of committed people, but fail once they move outside the orbit of those individuals.</td>
</tr>
<tr>
<td>Dismissing teachers with value-added scores in the bottom 3 percent who failed to improve</td>
<td>Success, <em>but . . .</em> in the pilot (done as a randomized controlled trial), extremely low-performing teachers (crosschecked by observers) were replaced with “average” new teachers. Students learned more the following term. But long-term, this was not a politically viable strategy, and it was hampered by teacher shortages.</td>
</tr>
<tr>
<td>Working with the government to encourage assistant principals to fill in as “semi-permanent substitutes” in unstaffed classrooms</td>
<td>Promising</td>
</tr>
<tr>
<td>Shorter teacher guides</td>
<td>Success</td>
</tr>
<tr>
<td>Reorganizing the school day to facilitate more on-task behavior from teachers</td>
<td>Success, Teachers are now spending up to twice as much of their workday teaching as they did before.</td>
</tr>
<tr>
<td>Once-a-week class in Uganda for 7th graders on how to set goals</td>
<td>Girls, <em>large success; boys, null effect</em>. We were not sure what to make of that, and Bridge is now testing in a different context, with tweaks, to see if effects hold up.</td>
</tr>
<tr>
<td>Cross-grade ability grouping for foundational literacy and numeracy</td>
<td>Failure in Liberia; <em>success</em> in Nigeria</td>
</tr>
<tr>
<td>Tutoring by phone during Covid</td>
<td>Success for grades 1 and 2 students in Kenya. Failure in another effort in grades 3, 5, and 6 (though phone calls increased student perceptions that teachers cared).</td>
</tr>
</tbody>
</table>
classroom interactions that spur imagination or critical thinking. Far from it. The incumbent method in the developing world is rote teaching (teachers talk, kids sit, occasionally repeat, and occasionally copy from the board).

Bridge used an approach similar to Piper’s, deploying teacher guides, training in using those guides, and affordable textbooks in classrooms that had often had nothing before.

Bridge also gave teachers electronic tablets on which to access the scripted lessons. The tablets doubled as a way to send data back to headquarters, which became a key part of the strategy.

And then the special sauce: observers like Imisi and Olu Adio in Nigeria, Gabe Davis in Liberia, Faith Karanja in Kenya. Hidden figures. It’s these field officers plus big data, working together, that help Bridge figure out which new ideas to try.

Bridge fails often (fast and slow) but ekes out and stacks up small, aligned wins in curriculum and other areas and walks away from ideas that don’t work out. (See Table 1 for examples of ideas that Bridge has tried, and to find out which ones have worked—and which have not.) The external evidence on Bridge suggests that the learning gains are real and large.

I believe future external evidence will bolster these claims, perhaps in a jaw-dropping way.

* * *

Imisi Olu-Joseph comes from a family of educators. She wanted to be a doctor, but her father, who runs schools himself, wanted her to be a teacher. “I majored in microbiology,” she says. “That was the closest thing to medicine he would allow. I started out by teaching in one of his schools. This field-team job is freedom for me. The motivation is seeing the improvements, little by little, and the boys and girls who make noticeable leaps from one visit to the next.”

Imisi was recently promoted at Bridge. She now leads all the school network’s field officers around the world. “I look for exceptionally intelligent people who can appreciate data and think deeply about complexity, how each thing affects another. Oh, and I need to avoid opinionated people, with strong preferences on instructional design. That type sees what they want to see.”

Mike Goldstein is an adviser to Bridge International and the founder of Match Education in Boston.