Using SABER: A Tool to Apply a Systems Approach for Better Education Results

Robin Horn
Director, Programmes, The Children's Investment Fund Foundation, and former Education Sector Manager for the World Bank's Human Development Network

So, I'm going to take off from Fernando's – one of his major points, and it came up in the questions. It's the issue of why is it that innovations are not being taken up by policymakers, by teachers, by schools, and scaled-up, especially if there's evidence that they're working.

Before I start my presentation, I'm going to step back to talk a few minutes about what it is that people who work in the international community – who are trying to build the quality of education systems to deliver better results, and to get outcomes of learning and some of these other skills that Fernando was pointing out before – not just literacy and numeracy, but all those 21st-century skills. How has the community been working on this? How have organizations like the World Bank or USAID, or the UK government, or NGO's Save the Children, local NGOs, Pratham in India for example.

These are organizations all around the world who are trying to get countries to improve their education systems, so that all children learn. All children become effective adults, and are able to go to higher education, and produce things for their economies, as well as for their families. So it's all the same objective. So the fact is, over the years, in all these communities that I just described, there is a kind of theory of change which is implicit – sometimes explicit for some organizations – which I'm going to describe now.

What I'll do is I'll describe this theory of change. And I'm going to talk about SABER, which is a tool – and I'll describe that – that helps address the issue of take-up, of ownership, of effective interventions and strategies. I'll talk a little bit about the methodology of that. Give an example about the teacher system and interventions to improve teaching, and ultimately learning. And I'll talk about all of this as an open data system. In a sense, that's where we link to technology – but in other ways, as well.

So what is this theory of change that drives billions and billions and billions of dollars of investment over the last 30, 40 years to help countries get better education outcomes and better results? The first step in this theory of change is this idea of identifying the determinants. This is usually done on the basis of theory. Fernando just gave some of the

theoretical justifications, the theoretical background for what would matter in developing effective strategies or interventions that could make a difference. It's also done on the basis of observation and experience. Where are things working well? What are the things that might matter there?

Let's pick those out, and let's do some studies. Let's do correlation analysis, regression, and more advanced research, more recently, to figure out what are the factors that make learning happen. How will computers help children learn? How will mobile devices help a teacher be more effective in the classroom?

First, you figure out what it is that makes that happen, and then you develop an intervention around that. And that's the second step. So you figure out this matters. Getting children to have mobile devices that are connected to the internet. This allows them to pull information out of the World Wide Web, and apply it in their classroom, maybe even in a more structured way.

So typical kinds of interventions that are developed might be – based on our research in the US, or at MIT or somewhere – getting children to have computers with learning objects on them, as Fernando was pointing out. Or, teachers using learning objects that that's going to make a difference. And this is how it would work. So you design an intervention. And you get that. You have a number of schools, or a number of institutions, or a number of universities where you try that. And then the expectation is we'll learn from that.

Down the line, we'll figure out what to do with that information that could help schools and universities everywhere in the world. So for example, you come up with some hypotheses, and you develop an intervention. And you test it out. By the way, some of these interventions – many of them – are implemented in settings, which are controlled by the researchers or by the designers of the implementation. For example, an intervention in Zambia to get students to have more textbooks, or an intervention in Guatemala to give parents more power in order to oversee and hold teachers accountable.

You have to make sure that all the conditions are right, and all the pieces of the experiment are in place – that your supervision is done properly, that you have control over the whole intervention – so that you can test whether it's working or not. You're able to figure that out. And there's a lot of science in doing that.

The next step is to evaluate that intervention. And the conclusion of that evaluation might be proof of concept. If you give computers to these students, in this way, and you measure these outcomes. And you make sure that there's a counterfactual. This group didn't get them – randomly assigned – this group got them. Great. And you get a rigorous evaluation, maybe done by J-PAL at MIT, or IPA at New Haven, or other organizations around the world that do rigorous controlled experiments with random controlled trials,

sometimes, or discontinuity of regressions. You come up with a rigorous finding. Fantastic.

You have a proof of concept – that if you do this in this certain way, in this particular period of time, with these controls, you get an increase, an effect. Maybe a sixth of a standard deviation of improvement. Wow. You get that effect. Next step is that you go, you take this finding, this rigorously evaluated finding – robust, with an experiment behind it – you go and take it, and you go to policymakers. Go to the minister. You go to the teachers union. You go to other stakeholders. And you say, we have it, we have a solution. This is what you do in this way. And you do that in that way. And you get the result – a fraction of a standard deviation in improvement, or even a half a standard deviation in those few exceptions where the intervention is phenomenal.

Of course the next step, which is the at the end of the theory of change, which is where religion comes in, in some sense. It's where faith comes in that, once you show this to the policymaker, once you show it to government, they say, "I'm going to do this. I'm going to not only take this thing that worked in seven schools or 20 schools, I'm going to take it and put it in all my 400,000 schools in the country. Because this is effective. And I will do it myself."

Well, the problem is that that last step doesn't happen. Or, I should say, it rarely happens. Sometimes it does. Sometimes it doesn't. 95% of the time it doesn't happen. So what's the point? What am I saying? For countries to take on and scale up interventions, first of all, they need to be involved in the process. They need to actually be implementers from the very beginning. Because it's not a proof of concept if it's not implemented by the owners, the implementers, themselves. All it is kind of an isolated controlled experiment that's done by an external group, with external financing, with external supervision And there's no way where that is a step where the next step is country ownership and large-scale expansion and delivery.

So what's missing? And that's what I'm going to be focusing on for the second part of my talk is the policy context, the enabling factors, the enabling conditions for interventions to be taken up, to be owned and delivered by the authorities or by other stakeholders who have the capability of reaching out beyond just a limited intervention. So that you want to see that intervention being viable in the context. You want to be able to know, from the very beginning, if it could be sustained in the system that exists. Dor absorption and expansion, ultimately, you need a set of system metrics to specify and measure the macro parameters that allow you to know, with some advanced knowledge, whether this intervention fits this context and can be absorbed.

Now I'm going to talk about SABER. Some people call it "Sub-air." For example, my colleague from Brazil would say "Sub-air," because it actually stands for System Approach for Better Education Results. That's what SABER stands for, but it also means knowledge in Spanish, Portuguese, and so on. First of all, what does SABER provide? It

provides an understanding of what matters most. That's a global "what matters most." It's not just what matters most in this context for this intervention, but what matters most to get interventions dealing with technology, for example, adopted anywhere in the world – rich, or poor, north or south, and so on. Across the key education system domains, the teacher subsystem, the finance subsystem – there are a bunch of different subsystems that make up the education system. What matters most in each of those subsystems?

It provides descriptive data on the policies. And I'll talk about that. And I'll show you what I mean by that. It assesses. It evaluates. It judges what the policies are, whether they're good policies, and ultimately whether they're being implemented. Then there's country ownership and sign-off. They are part of the process. Government is owning this process for measuring system policies. And, finally, it puts it all out there, so countries can learn from one another. So researchers can learn from what countries are doing. And you could learn from the real-world practice.

So the theory of change with SABER in it is a little different than the theory of change that I walked through before. Because it says that these interventions – that are designed in order to test a certain hypothesis, and then evaluate rigorously – have to be designed in a way that are consistent with the policies that are in place in a given country. For example – and the SABER looks at the policies.

For example, if you're working with an intervention that requires that you improve the quality of the system that supports teachers to use technology – I'll stay with these technology examples because it's relevant for this particular audience. But let's say that your goal is that you have a hypothesis. You have a set of theoretical reasons to assess, to design and test an intervention that would use the existing teacher system – to have those teacher's supervisors who are out there in the districts, let's say, to improve the use of technology in the schools, to work with principals, for example.

You have a whole model. It worked great in Berkeley, California. Fantastic. It's beautiful. The system works. The teachers were getting that support, and so on. So let's bring it to Bihar, India. Everything's fine. Except let's say there's no teacher support system in place. There's no policy that says there should be a teacher support system in place. So the implementers say, OK. Well, let's create a little mini teacher support system in this district in Bihar. Then once we get that working, we'll show the state government that, if they only did this – created this whole system, passed laws to put this policy into place – great, then it would work.

So that's an example where you have policies, and you have interventions, and they don't fit. You have an intervention that doesn't fit with that particular policy. So what you need to do is figure out what are the policies, what is the policy framework, what types of reforms, what types of interventions could this particular country support, given its policy framework? Then you can figure out – with the counterparts, with the country – what to do, how to do it.

So what they did with SABER – I call it "Sub-air," sometimes SABER – is look at the policies that, based on the billions of dollars that the World Bank has provided to countries, and the huge amount of research that many of you and others have done all around the world have come to look at.

What are the systems, the subsystems of an education system, whose policies are important for getting good results, and whose policies need to be understood? The policy framework needs to be understood better, so that you can design interventions that could eventually be taken up by the government, by the stakeholders in place. So what we have here, for example, we have policies around learning standards. What is it that students should learn to be able to do? What are the policies? Are they standards-based? Are they just content coverage? And we're able to say some policies make more sense – from the research, from observation and from pure research – than others.

Some of the weaker policy in learning standards would be just a set of coverage topics. Teachers are expected to go through all the topics from the beginning of the year until the end of the year, irrespective of if any of the students are learning anything. So that's not a great policy in learning standards. The other extreme might be standards-based, where there are tools to assess each student, whether that student has mastered that standard. That's more advanced.

So one is kind of latent, or undeveloped. The other extreme is advanced. And you start assessing that. That's learning standards. Financing, teachers, ICT. World Bank is developing policy frameworks for ICT, for health and school feeding, workforce development, ECD, and so on. So all of these things – I'll show you at the end how you get access to all this information.

What are the steps? I'll take you through the steps for SABER. One step is to carry out a state of the evidence review, could be state of the art, but of the evidence. What do we know? What does the evidence tell us? What is the research? Good research – what does it tell us about what matters in developing good learning standards? What matters most in good teacher policy? In addition, there's a lot of research that doesn't exist as Fernando was just saying. There's huge amounts of things that we don't know. There is a lack of research in policy frameworks, and there's also a lack of research in interventions.

So you need to also learn from countries that are performing well. For example, this country seems to be performing very well, according to some international tests and other measures on learning standards, on their learning standards system. Let's see what they're doing, what their standards are. And then we could infer from that what seems to be good learning standard system. Ultimately, I'll describe that you can test that and evaluate that in practice over time.

Next step is we – the team working on that particular subsystem, let's say learning standards or teachers – develops a kind of state-of-the-art book on what matters most in

good policy for that topic. We'll call it "what matters," and in that are the metrics for measuring policy.

Until now, until SABER, there were no metrics. There was no way of comparing countries' policies with one another. Each researcher goes to a country, or a state, or a province and brings his or own language, his or her own tools, applies it, and comes up with a report. Maybe gets it published here. Maybe gets it published there. But they're not comparable. So one of the things that's necessary, in order to have consistent knowledge and consistent measures of policies, is to develop these indicators, these metrics. Define them in a crystal clear way, and then apply them in different contexts in different countries.

Then also there's a scoring rubric. So based on the data that's collected, what is OK? What seems to be – from the evidence – what's good? What's great? We call it Latent, Emerging, and up through Established and Advanced – the kind of stages of development. You could also say one, two, three, four. You could also say bad, good, and so on. But we came up with these terms because countries seem to accept them more readily.

Then there are these instruments that collect this policy information. All of this is publicly available, by the way. And I'll show you how you could get that. Analysis of the data. Rating of results. Countries then validate the findings. The ministries of education or finance, depending on the topic, will look at this and say, yeah, this is accurate. This does reflect what we have in place, in terms of our policies. You can go ahead and publish that in your open data.

That's not an easy step, as you can imagine. But countries that are interested in learning what their neighbors are doing, or what their comparative countries are doing, they can't say, I want to see everybody's, but no one can see mine. Because if everybody had that approach, then no one sees anybody's data. So they accept it. They buy into it, just like they do with PISA, those of you who know PISA or TIMSS. And then, this is put out in Open Data, on a knowledge platform database.

Here's what I was talking about – levels of development on the rubric, from Latent, to Emerging, to Established, and Advanced – to show whether this is a policy that's robust or a policy that's weak and needs to be strengthened. And countries could use that to target a particular policy area and develop it further.

So now if I want to give an example of one of the policy areas – one of the domains on teacher policies. In terms of the context for this, until SABER came along, there was scarce data on teacher policies, per se, for developing countries. There were a lot of studies of interventions, a lot of theories out there. But in terms of the policies, there was very little. And there was very uneven evidence of what works and limited guidance on policy decisions.

So that was the problem that SABER teachers tried to address, and then pulling together global data on teacher policies, comparative analysis, opportunities to learn from education systems around the world, and you end up with basically this chart. If you could read it, it says these are the eight policies that, when implemented well-- not necessarily all of them well, in the same place – but when these policies are implemented well, results are better. Outcomes are better.

One of them is setting clear expectations for teachers. Attracting the best qualified people into teaching. Preparing teachers with useful training and experience. I'll stop here. You'll be able to see this later. This all goes on the web. I'll be focusing for a second, with the next slide, on this goal. What does it mean to have a good policy on preparing teachers with useful training and experience? Basically there are two main categories. In the end, it seems very simple. But this is something, looking across the world, across hundreds of studies – what studies exist – and shows that there are basically two factors that you could look at across the world.

Are there are minimum standards for pre-service teaching education programs? And to what extent are teacher entrants expected to be familiar with the classroom? When both of these are in place, then you could say that, at least, it's a reasonably good policy for preparing teachers for effective instruction. So what is the minimum level of education required to become a teacher, for example? And we apply that to primary school teachers and secondary school teachers. And your teacher entrants have opportunities to learn from practice. Again, we look at this separately on these two types. How much experience is needed? And sometimes it might be, in some countries, the experience is zero. You just get theoretical training, and you're off and teaching. They would get Latent on that. In other countries, they have to stop practice a year under supervision. That's high quality. That's at the high end. And that's where you'd have Advanced.

So now, in my closing, I will show you some examples of the website that the World Bank has now, as it exists, recognizing that in a couple of months it will change. They've been investing to make it into a more practical, usable website, so that all researchers, all politicians and policymakers can access these data and use it when they're designing interventions or trying to improve their systems. By the way, there are six of these "What Matters" reports all available, 12 domains in design or under pilot, from teachers to ICT, and so on. But 100 countries have at least one domain, usually multiple domains. And there are a bunch of country reports that have been released by countries.

So I don't know if you can see this. This is from the website. I'll give you the address in a second. This shows, at least on the right-hand side – you scroll down. There are 90 countries there, including some provinces and states in some countries. And the colors indicate – black means the data is already approved by government for release on the Open Data platform; dark blue means it's completed, but not yet signed off by government; and light blue means it's still in the process of data collection or analysis.

Here's an example of the page on teachers. If you look at this, here are the goals on the left. So here, for example, the third one is preparing teachers with useful training and experience. Then you look across – you could scroll across all the maybe 50 or 60 countries that have done this. For example, on the first one, let's say, setting clear expectations for teachers. Egypt on that particular metric, on that indicator, received a score of Advanced, whereas Jamaica, the second one, received a score of Emerging. So then, when the Jamaican authorities look at this, and they say, what do you mean? What's going on in Cuba? What's going on in Brazil?

They're able to compare themselves with other countries that are of interest to them, and say, why is it that Saint Lucia has Advanced in terms of attracting good teachers? And then the report will – I'll show you the report in a second – that will take them through what it is that they're doing, what they could do to improve their policies, and so on. So you could see that. Here's an example of a snapshot of the Country Report that the country gets that they sign off on, that's in the public domain. Here's the front page, which shows, in each of those different goals, what the status is. For example, this is Jamaica. You saw a second ago that it's at Emerging for setting clear expectations for teachers. There's the score there. There's a little text there. And then inside, in the report, you have one page per goal, so it's very accessible.

That's the other thing. It's really important. Information needs to be accessible for the policymakers to understand it, so they could use it. And here's a snapshot of a page. I'm not going to read it, but this is the page on matching teachers' skills with students' needs. You have one page. And then on the right, you see a chart, where Jamaica said, well, these are the countries that we're interested in. We're interested in Mexico, but we're also interested in what the advanced countries are doing. And so we agreed with them. OK, let's show Singapore, Japan, and they could compare themselves with those countries on these measures here.

So in closing, this is the website. I'll leave it up during the question and answer period. We have not quite two minutes for maybe a couple of questions. And then later on, maybe we'll have more time. Thank you.