ABCDE 2025

Thursday, July 24th 2025
Washington, DC

EDUCATION POLICY

Eeshani Kandpal: All right, I think we're going to get started. Remarkably a minute early. So if you're David Evans, please have a seat. All right, this session is titled Education Policy, and it is moderated by Deon Filmer, who is the Director of the Research Department at the World Bank. This will go for an hour and a half, and then we're going to go straight to a cocktail reception after that. So I will not be back up after this session, but see you all tomorrow.

Deon Filmer: Thanks. Great. It's actually a better turnout than I feared for the four o'clock session. So I do not know that a session on education really needs a lot of motivation. But I do want to tie it a little bit to the conference. So this morning, Rachel kicked us off by saying there's no sustainable, I'm paraphrasing, no sustainable growth without human capital. So obviously human capital is central to the development part of this. But then on the first day, Masood, in response to, I don't know if it was a question in his talk where he said: What should we do? What should countries do? What should agencies, multilateral agencies like the World Bank do in the context of declining multilateralism? And his answer was: invest in the basics of growth, including human capital and in particular, education. So there we have some motivation coming from within our conference for this session. So as I said, the session is going to be focused on education policy. This morning, we heard a lot about service delivery. This session takes a slightly different angle. We're going to start with a a big picture story about the global distribution of skills.

And then we're going to go into three specific constraints to learning, and they're not the typical ones we think about. So the idea is to simulate a little bit of new thinking in this area. So the speakers are going to be Dev Patel, who's going to talk to us about the global distribution of skills. He's the Prize Fellow in Economics, History, and Politics at the Center for History and Economics at Harvard University. He's going to be followed by Christina Brown, who is going to talk to us about cognitive endurance. And if you don't know what that is, she's going to explain it. She's an assistant professor at the Department of Economics at the University of Chicago. Then we're going to hear from Lee Crawfurd, who's going to talk to us about lead exposure and how that impacts learning. He's a Senior Research Fellow at the Center for Global Development. And then we're going to hear from Gabriela Smarelli, who's going to talk to us about violence in schools. She's a Senior Research Associate in the Global Education Program at the Center for Global Development. So let me invite, without further ado, Dev.

Dev Patel: Thank you all so much for being here, and I'm delighted to present this joint work with the brilliant Justin Sandefur, who's right here. This project is entitled the Rosetta Stone for Human Capital. It's about trying to answer very basic questions like this one: Who can read better, a fourth grader in Niger or Peru? While questions like this may seem very simple, they're actually deceptively difficult to answer. The reason why is that we lack a global standardized test. One way of visualizing that is with this map. This is coverage of the primary school standardized test exams with the largest coverage, the TIMSS and PIRLS exams. As you can see from this map, most of the world is actually excluded from the sampling frame, including all low-income countries. Parts of the developing world that are included in standardized tests tend to participate in these regional exams, like the ones shown here. And they're strikingly non-overlapping and also quite sparse. So what this project is trying to do is try to answer questions like, if we took a student from Brazil who took LLECE, and then we took a student from the United States who took TIMSS, how can we actually compare those test scores?

For example, students in both of these countries have participated in these standardized tests, but the test scales on which these scores are measured are completely different. How can we compare, for example, a 600 for a student in Brazil to a 500 for a student in DC? That's going to be the question we're going to tackle today. This paper has two parts. I'll just give you a high-level overview of where we're going. The first part is trying to understand how we can actually translate test scores from one test to another when they're completely separate on these different scales. So from a high

level, what we do is we take different tests. So for example, you can imagine we take a yellow test, a green test, a red test, and a blue test. We build a hybrid exam that takes questions from each of these tests and put them together in one booklet. Then we give that booklet to some students. And when we grade that test score for that student, we do it on each original test score scale. We grade that student's performance on the yellow scale, the green scale, the red scale, and the blue scale.

We do this a bunch of times for many, many students. We can then actually compare for any other test score from a student who only took one of the original exams of what they would be like on another exam. We can create this Rosetta Stone translation from one exam to another. Then once we have that, we can go and find a bunch of other students who may have only taken one of these tests, pretend like that green student actually took the yellow exam, and put them all on the same scale. Once we've done this methodology, we're going to actually implement this to try and link four of the world's largest standardized tests for primary school students. So this is an important component of, for example, the SDGs. We'll do this to actually translate test scores for about 600,000 kids from 80 countries for both reading and math. And what I want to bring to this panel is some basic statistics about the distribution of learning around the world. So we're going to start with four main facts. The first is that where a child lives turns out to be much more important than how much money their household earns in terms of predicting their test score.

In other words, if I took two students who have the same household income, but who happen to live in very different countries, it turns out that that is going to matter a lot more, which country you live in, as opposed to how rich your household is. The second factor we're going to find is that the relationship between your household's income and your test score is going to be steeper in countries that are more unequal. This is going to be very reminiscent of the types of other findings on inequality and the so-called the Great Gatsby Curve that you might be familiar with. The third fact is that we can look across gender, and at a household level, if we look at your household income, we see that no matter what household income you're in, whatever bin you're in, it turns out that girls tend to do better than boys in reading. However, that's not true for math. For math, only the poorest households are going to see that the girls do better than boys. In fact, we're going to see a reversal as households get richer. And then the fourth fact we're going to find is that if you look at the test score gap between public schools and private schools, we're going to see that the difference in scores is going to be bigger in countries that have more income inequality.

I'm going to dive into a little bit more detail now on all of these facts and our methodology. First, I'm going to start with just providing a very high-level overview of the method that we developed and how we implemented that using field work we conducted in India and in Florida. Imagine we took our four different color tests. So we started with our four color tests. One thing you might do is just give each of those tests to students, and then you would grade those tests. Now, the problem with that is that tests are really long, and no student really wants to sit through that many tests, let alone one. So that's the first obstacle. The second obstacle is that a lot of the tests that we use as researchers are not actually publicly available. So even if we wanted to do that, those testing companies are quite protective of their standardized tests. And so they only release a small number of questions. So the way we try to overcome this is we actually build an exam that takes test questions from each of the original source tests and puts them all together in a new test booklet.

This is going to be what I'm going to call our hybrid exam. So then from a high level, what we can do is take this hybrid exam and give it to students. We did this both in Bihar, India, and in Florida, in the United States. For each student, the core methodological part of our paper is to use statistical methods from psychometrics from this Item Response Theory that's used to grade the original test. For each student who takes our hybrid exam, we're going to grade their performance as if it were the original full test from the source exams. This is the exact same type of method that's used by the SATs to compare test scores over time or across students, and we're just going to apply that to this setting. We're going to do that not just for one student, but for many, many students. And what

we're going to find is that for any given person on that same day that they took this exam, we can assume they had the same ability. And so we can say, okay, they scored a 486 on the red test, that must be equivalent to a 312 on the blue test.

We can do this for a bunch of students. And that means that across the ability distribution, we should be able to compare students and to do these mappings from one test to another. This is the core method we've developed. We're going to do that and apply that to four big standardized tests. The TIMSS and PIRLS exams, which cover the OECD; LLECE, which covers many Latin American countries; and then PASEC from West Africa. We're going to develop these hybrid exams, pooling questions from these different source tests and administer them in both Florida and Bihar. This is important because this actually gives us a pretty big coverage across the skill distribution. For the rest of today's talk, we'll really focus on some of the math results that we find. So once we do this, what we can come up with at the end of the day are test conversion functions that look like this. So just to walk you through this, On the X-axis of these graphs are the scores on the TIMSS exam, the OECD exam. And because we graded for each student their performance on another exam, so on the Y-axis in the graph on the left is the West African exam, and the Y-axis on the graph on the right is Latin American exam, we can map from one test to another.

What you can see is that we can just convert, put in a test score on one axis, and you can draw a line and figure out where it would be on the other one. These are the types of test score conversion functions we can estimate and then apply around the world. What that allows us to do is, for example, take those original test score distributions from Brazil and the United States and now put them on the same X-axis. We can see exactly how those students now line up. So once we've done that, we're going to now talk through a little bit more of the different types of facts we can learn from this type of big learning data once we've harmonized test scores around the world. I'm going to, for the rest of the talk, focus on test scores that are in the units of that TIMMS and PIRLS OECD exam. So you should think of 100 points there as one standard deviation in that OECD units. Okay, so fact one that we're going to find is that where students live turns out to be much more important than their household income. I'm going to visualize that in a simple graph

Here on the X-axis is the household income of the student who took the exam. I've drawn dotted vertical lines here the World Bank Income Classifications. The Y-axis of this graph is the student's test score on that TIMSS and PIRLS OECD exam. What you can see is that, first, there's a really big difference in math and reading ability as you move up the income distribution. Richer students are performing significantly better than poorer students. When we talk about the learning crisis around the world, this is showing a lot of data towards that fact. But one thing that this binned scatter plot masks is that if you then try to do the same thing separately by country, you see there's a huge amount of variation depending on where students live. Now what we've done is in each shade of blue here, the lighter shades are high-income countries and the darker shades are lower-income countries. What you can see is if you just take a vertical line and draw it somewhere on this graph. So you're holding fixed the household income of a student. It turns out that what country you live in is going to be a huge determinant of what your test score is.

And it turns out that that variation across countries is going to be much more important in our data than the variation within countries across household income. So that's fact one. The fact two is about how incomes and test scores relate to one another. I'm going to plot in this graph two very simple statistics. On the X-axis is the Gini coefficient that we're all familiar with about how unequal the income distribution is within a country. The Y-axis is the relationship between a country's student's math score and their household income within that country. You should think of a separate parameter for every country. For a country like Finland, which is a very unequal country in terms of its income distribution, you see that the relationship between math score and income is also very low. In other words, if I increase my income quite a bit in Finland, my test score does not go up very much. By contrast, countries like Brazil, that's a lot more unequal. The income side is also going to

have a much steeper relationship. So the richer the students are going to do significantly better in terms of their math score.

And this tends to hold across countries, too. So we can now plot all the countries in our data set and we see this same relationship. The third fact I want to show you today is looking at the gender gap in test scores. Here, what we're going to do is that same graph I showed you before, where the X-axis is the household income and the Y-axis now is going to be the ratio of test scores for girls to boys. If it's above that horizontal dotted line, that means girls are going to do better. When we look at reading, we see that across the whole household income distribution, girls are actually outperforming boys in terms of reading. When we look at math, we see a different story, however. Only in the poorest households are girls outperforming boys in math, and actually, you see that relationship reverse, as households get richer. And then the last fact I want to talk to you about today is about private schooling. This is obviously a huge phenomenon around the world. One thing we can do with our test score data is just compare the test scores of students in private schools to those in public schools.

So here in Colombia, for example, private school students tend to score about 50 points. This is like 0.5 standard deviations higher than students in public schools. We can do this across all of the countries in our data. You can see that there's a lot of variation across countries in the differences between private and public schools in the raw test scores. Of course, there's differences in the types of students that go to private and public schools here. This is not a causal impact. But if we put in a bunch of controls for things like gender and wealth, things change, but they don't change a huge amount. There's still quite a bit of variation. One thing we can do then is see how this varies with inequality in a country. So now on the X-axis is how unequal a country is in terms of its income, and the Y-axis is that private school premium. What you see is a positive relationship here. So the more unequal a country is in terms of income, you also see that there's more of a gap between the private and the public school students. Now, it’s very consistent with, for example, in unequal countries, if there's more socioeconomic segregation across schools, that's going to manifest itself in the same way.

So I'm going to just conclude now by saying this paper was about trying to compare test scores across countries, even when those countries participate in different standardized tests. What we do is develop a method to link those exams together. There's a lot of learning inequality, and that is certainly popping out in our data. Very excited for the rest of the panel. Thank you so much for your time.

Deon Filmer: Thanks, Dev. Next up, we have Christina.

Christina Brown: Great. Thank you so much for having me. Delighted to present this research, which is joint with Supreet Kaur, Geeta Kingdon, and Heather Schofield. So I think I don't need to convince this room that we have lots and lots of evidence on the effect of schooling on many long-run outcomes for individuals, everything from income, health, criminal behavior, social outcomes. We think schooling is really, really important for lots of long-run outcomes. And we have some evidence on what are the specific channels that might contribute to those long-run outcomes. But I would say they fall into these two broad categories. There's some research that focuses on this idea that it's the specific curriculum, the academic content, the literacy and numeracy skills which are contributing to these long run outcomes. And then there's this other idea that schooling probably does lots of other things, too, in terms of building people's cognitive capacity. It might influence their ability to improve their working memory or their attention skills. And in that second piece, we have much less evidence. We really don't have a lot of great measurements about those fuzzier cognitive skills as compared to all of the really rich evidence that we just saw on measuring concrete literacy and numeracy skills.

And so what we're going to try to do in this paper is really focus on one specific aspect of schooling, in particular, when we're going to think about the fact that in most schools or most good schools,

you're spending a large part of your day on effortful thinking for an extended period of time. That's the experience of what it means to go to school. And then we're going to look at the effect of that type of schooling on a particular cognitive capacity, which we're going to call cognitive endurance. So what do I mean by cognitive endurance? We're going to define this as the ability to sustain a performance on a cognitively challenging task over time. So this is, how much are you learning from this talk in the first minute versus the 15th minute, or at the end of this panel versus the beginning of the panel, how much are you taking in? That's really going to be our measure here. And there's been lots of evidence showing that this has broad relevance for a lot of different settings. We've seen that in the workplace, people tend to make more mistakes later on in the day, even in this situation of voting, further down the ballot, people actually make less active, effortful decisions, and they tend to rely more on defaults if a ballot item appears later down.

But interestingly, this really varies by overall income of the population. So it's not just that everyone gets cognitively fatigued, but that this might be particularly worse for certain populations. Okay. Oh, I realized that the slides are slightly old, so I'll have to improvise a little bit on some of them. But what I'm showing you here is that not only do we see this cognitive fatigue happening in the workplace, it's also a really common feature of standardized tests. So in the TIMSS test, again, more TIMSS data, building off of Dev's work, we tend to see that over the length of that exam, so this is within a single subject, which is about a 30 to 40 minute exam, this is for fourth graders, that students tend to perform worse over the length of that task. So relative to an identical question item, because we can actually control for question items, there's variation in the ordering of questions that kids get in the TIMSS test. For an identical question item, kids are performing worse when they happen to have gotten it as the last item versus the first item. But this really varies by income level. So this fatigue that students are facing is substantially worse if they're coming from a low income country.

And there could be lots of explanations for that. It could be that, for example, there are differences in nutrition. Maybe students didn't have time to finish the test. There could be lots of drivers of that fact. But we're going to focus on one in particular, which is that maybe the type of schooling that kids experience in low income settings is such that in a high income classroom, they're getting opportunities to practice training this cognitive endurance muscle, right? They have the types of instruction where they're getting to practice attending to a task over an extended period of time. And so that way they end up being able to deploy that skill in lots of settings. And in particular, you can see this when you look at classroom observation data as well as teacher self-report. This graph is showing you on the X-axis is income per capita or the % of students coming from disadvantaged groups. And the fraction of time that students spend as part of the school day on individual-focused practice. So in low-income settings, we typically think of classrooms as being much more teacher-led, that students are listening, copying down what they're hearing from their teacher, but not spending a lot of time on individual-focused practice, whereas in high-income countries, they're spending a much larger fraction of their day on that type of activity.

So what are we going to do? We're going to conduct an experiment with a series of private schools in Lakhna, India. These are going to be low cost private schools, and they're going to have all of those same features that I just talked about, where most of the instruction that students are going to be experiencing is teacher-led lecture, where students are passively listening and maybe copying down notes, but they're only spending maybe a few minutes during each period of the day on concentrated cognitively challenging work. And so what we're going to do is try to come in and say, okay, let's try to get it so that a larger fraction of their school day is spent on extended practice on some cognitively challenging activity. We don't really care what the activity is. We just want them to spend more extended time on some thinking activity. And in particular, we need this to hold their attention for 20 minutes or so. This is from the length of their periods in these classes. And so our idea was to basically introduce a series of different very basic apps, not gamified, very stripped down, somewhat boring, but that were dynamically adaptive The idea being that if we wanted

something that just barely is holding these kids' attention, so it's stretching that attention muscle, the content can't be too hard that students are tuning out, and it can't be so easy that they're tuning out for that reason.

And so what we did exactly was that coming into their existing elective period, so these are 30 minute periods that they have a couple of times a week, we randomized students to either receive the math practice tablet application or the games practice one, relative to the status quo study hall period. So in the math and the games group, they're getting a tablet during this 30 minutes period. They're working on either math problems or in the games group, it's like puzzles, mazes, apps that we specifically chose to be, again, very boring to a certain extent, but also free of any academic content. We wanted them to be doing something cognitively challenging, but that didn't have any specific literacy or numeracy requirements. And that's going to be compared to this status quo where students basically get a study hall period. The teacher spends most of that time doing grading or other activities, and they write a few problems up on the chalkboard. Students work on those and then usually talk to their friends or do some other activity. And so in practice, what this ends up meaning is that the treatment group is going to get an additional about 15 hours of focused practice over the six month intervention period relative to the control group.

So they're getting 15-ish additional hours spread out over these different periods of time. So what do we find? Well, what we were fundamentally interested in was if you get them to focus it practicing on math problems or on puzzles and games, is this a general skill that then they can deploy in lots of different settings? Or is this a very specific skill? That if I practice attending to math problems, I can pay great attention to math problems, but that's it. Is this an underlying general skill that can be transferred to different settings? So what we did was we tested students in three different domains, some that were more or less related to the activity that they had been doing in their elective classes. And we randomized the question order so we can look at — our outcome of interest here is not just how well you perform on this listening comprehension task, which is one of them, but how much worse do you do at the end of the test relative to the beginning? So this first graph is showing you for the control group. Sorry, it's a little bit small, the labels here.

What students' performance looks like over time. And for the dotted line here, the control group, we see that students are doing about 3% points worse on an identical question item when it appears at the end of the test versus the beginning. And this was a listening comprehension test that kids took basically listening to headphones, they listen to a short story, and then have some understanding check questions that they have to complete based on... It was all pictorial, so there were no written literacy requirements to complete this. But it's a boring story that they're listening to. It's boring audio that they have to follow along to. And so really we're trying to capture, can you sit and pay attention, similar to maybe this setting currently, can you sit and pay attention to something where it might not be doing a great job of necessarily holding your attention. And so then the gray line was for the control group, and the blue and orange lines here are for the treatment group. And what this is showing you is, the first interesting pattern is that at the beginning of the exam, we actually don't find any difference in performance.

And that's what we would expect, right? Neither of the treatments were doing anything that would improve your underlying listening comprehension skills. It wasn't building those particular skills. But what it was allowing you to do was practice focusing on something for an extended period of time. And so we see by the end of the task, now both treatment groups are doing significantly better than the control group. They still suffer a decent amount of cognitive fatigue over the task, but it's substantially less than the control group. Similarly for the Ravens test and the math test, we see a somewhat similar pattern, though a little bit more noisy. So for the Ravens test, we see a little bit of a difference between the control and the treatment groups, though it actually isn't statistically significant because this one zoomed in a little bit. But we do see by the end of the test, you can see that's a little bit of a flatter slope, essentially for that blue and orange line compared to the gray

line, which has a steeper decline over the length of the task. On the math test, you also see this really interesting pattern. I don't know if you can see closely enough what's happening here.

But at the beginning of the test, we actually see that the games group, which is the orange line, is doing slightly worse than the control group. And that is also what we would expect, because if you remember, the control group is doing some math problems. They have those initial few problems that they're working on at the beginning of the study hall period, and then they're talking to their friends for the rest of the 30 minutes. Whereas in the games group, they're not doing any math-related content. So it's maybe not too surprising that in terms of their initial content knowledge, the control group would be doing a little bit better. But over the length of the task, and this was our longest task that we had, this was a 30-minute task, you see this real stark decline over the length of the task, and in particular, a slightly flatter slope for those who were in the treatment groups. And so then what we try to do in the rest of the paper is try to really understand: "Was this fundamentally about us training attention skills?" Or we were introducing tablets in the schools, maybe this was doing something else.

Maybe kids were more motivated, or they were showing up more for school, or they just felt special because they were selected for the treatment. So we tried to do a few things to figure out which of these explanations are driving the results. Since I'm short on time, I'll go through this a tad quickly. But the key thing to take away is that for a lot of the other channels that you might have in the back of your mind that could be impacting potentially by this treatment, we expect there also to be an effect at the beginning of the test. The fact that we're only really seeing positive effects later on into the 10 or 20 minute or 30 minute task says something about attention potentially being the channel. We also are able to measure some other channels and don't seem to find evidence that that's what's driving results. And then the other thing we try to do to narrow down these channels is that there are some different measurement tools from the psychology literature of how to measure sustained attention, which we think of as really analogous to this idea of cognitive endurance.

And we find positive evidence for two out of these three measures. And then the last thing that we do is see, okay, we've gotten kids to pay better attention to our specific tasks. Does this have any real effect on other outcomes we might care about? And we find that, yes, actually having kids focus on essentially puzzles and games, we actually find improvements in their English, Hindi, and math grades at the end of the school year. And quite interestingly, you might think, well, this is just coming from the fact that because they can pay better attention on tests, they're not fatiguing as much on their end of term exams. And all of this great improvement is really because they're focused better on those tests. When you do a back of the envelope calculation, that would only explain about a 0.02, 0.03 standard deviation increase in their exam scores. And so actually, we think most of these great effects are coming from the fact that if they can pay better attention in class, they're actually learning more. And so that's why we're seeing these improvements in grades, it's because this ability to attend is now allowing them to learn more content as well.

And then the last thing we do is try to say, okay, this was maybe a little bit of a niche experiment that we did where we came in and introduced a particular type of elective period. If kids are just exposed to traditional schooling, does this also improve their cognitive endurance? And so we use a fuzzy regression discontinuity design, comparing very similar age peers who were on either side of a birthday cutoff. And see that for a nine-year-old who is either in fourth grade or fifth grade, depending on their birthday, those who have an additional year of schooling have significantly more cognitive endurance, specifically if they attend schools where there's more time spent on this individualized, focused practice. So I'll just leave you with a few takeaways. I think the goal of this paper was really to think about [that] there's lots of things that we think schooling might do in terms of improving students' skills. We're focusing on one in particular, one in particular that we think from the data looks like schools aren't spending as much time on low-income settings. And so our hope is that not only we can show that there's ways to measure these other skills in a relatively feasible

way, if you have an assessment where there's randomized question order, you can look at declines in performance over time.

And so we really encourage people to start thinking about what are these other cognitive capacities in terms of measuring working memory or other measures of cognitive and noncognitive outcomes, and see what are the underlying aspects of teaching that could contribute to those skills. So I'll leave it at that. Thank you so much.

Deon Filmer: Thanks, Christina. That was fascinating. Up next is Lee Crawfurd.

Lee Crawfurd: Thank you very much. And thank you all for being here. So I'm going to talk about lead exposure, which is, I think still attention has been really growing in this space in the last couple of years, but I think it's still a massively neglected space. And so in this paper, we're trying to look at what this means for education in poor countries. And this is a joint paper with colleagues at CGD that was just published in this month's edition of the World Bank Research Observer. So if there are any copies in the building, I'd love to get one before I leave. So the fact that lead is correlated with IQ is really not a new finding. This paper is from 1979 in the New England Journal of Medicine. But IQ points don’t mean a huge amount to most of us development economists who work on education. We tend to think about standardized effect sizes. And so we came to this with the question, how big of a deal is that really for education in our countries? What's the effect size? And does it carry across from IQ into the literacy and numeracy skills that we care about as well?

And so this graph really frames the puzzle. So on the left-hand side, most children in high-income countries can now read a sentence by age 10. It's not most children in low and middle-income countries, where it's 29%, we think, around there, can read a sentence by age 10. And at the same time, lead is nearly a solved problem in high-income countries, just 2% of children with elevated blood lead levels. But nearly half of children in low and low mid-income countries also have elevated blood levels. So we know that there's some relationship between lead exposure and cognitive skills, but we don't know how much of this gap is lead explaining. And so that's what we set out to do with this paper. And so we started by doing a systematic review of a literature out there, looking at all of the observational studies that look at a correlation between blood lead in a child and their test scores, either on IQ or on reading or numeracy. We make some adjustments to all of these studies to try and make them comparable in the first place, which they aren't necessarily. Try to adjust for unobserved confounding because most of these are observational studies, and do what we can to show you that this is a causal relationship.

I adjust for publication bias in the literature, of which it does seem as some, but it doesn't fundamentally ruin this relationship. Then finally, we do the simulation. We go to each country and we look at the average blood lead level. Then we have this parameter from our meta-analysis showing the relationship between blood lead and learning and say, okay, so what would happen if we could reduce blood lead levels in low and middle-income countries to the levels of lead in the US, which is very low. What would that do to learning? So this is our meta-analysis. We find 286 estimates from 47 studies. Most of these studies are on IQ, 40 studies. The average effect size there is 0.22 standard deviations for each log unit. That's roughly a doubling or halving in blood lead, so a reasonable effect size. We also found a reasonable number of studies that measured impacts on reading and maths. And what's striking is that the average effect size there is almost identical across views. So we think of IQ as maybe being something different, but the correlation with blood lead at least, is very similar for IQ as it is for reading and numeracy, these foundational skills that we tend to care about as economists, focused on human capital.

So that's a simple unadjusted correlation. And obviously, we don't just want to take that. So we tried to adjust that. With apologies to Xavier Sala-i-Martin. We ran not quite two million regressions, but we ran every possible meta-analytic regression adjusting for different features of these studies. So a lot of these studies don't adjust for parental income or parental IQ, parental education, which you really should. And so we adjust in the meta-analytic regression for the studies that do and don't

make that adjustment. Our ideal study would be not in a high-income country and would be measuring reading or mathematics. And so essentially, when we go from our completely unadjusted estimate, which is the red dot here, which is the 0.22 standard deviation. Our preferred estimate is over on the right, the blue dots, and it roughly halves, so it's 0.12 standard deviations. But it's fairly arbitrary, and so you can see all the different choices you could have made in this meta-analytic regression. There are some differences in the estimate, but it's around the same ballpark, regardless of which choice you make there. So we proceed with our 0.12 standard deviation reduction in learning for a rough doubling in blood lead levels.

Now, it's a room of economists, so are these correlations really causal? This is an obvious question. I think we have a number of reasons to think they are. The first being simply that these correlations are relatively robust to including different controls and also the coefficient stability methods for unobserved confounders as well. Second, there's a really clear theory of change here. There's a clear, well-understood biological mechanism by which lead prevents the brain making connections. Third, perhaps most persuasive is this increasing crop of natural experiments in economics, which do claim to show causal effects. And by and large, the effect when using a natural experiment is larger than when just looking at the observational. That's true, I think, for all five or so studies which allow you to show both an OLS estimate and an instrumental variable estimate. They all show bigger effects with the instrumental variable estimate. And finally, perhaps ethically questionable, but there are some direct experiments with treating animals with lead that do show causal effects on the cognition of monkeys and rats. So I think it's quite a strong case. There's a causal relationship here. And so what does this do? Here's our simulation. We take for each country the harmonized learning outcome, which is a similar approach, not as fine-grained as Dev's approach to putting different countries on the same learning scale.

And on this scale, it's roughly 500 points the rich country average. And this set of developing countries is around 100 points, 150 points below that. And for each of these countries, we say they have different levels of blood lead in each country. And so the effects of eliminating that lead would be different for each country. But it ranges between 20 or 30 points on this scale that they could potentially gain if they could completely eliminate blood lead. And so that's enough to reduce 21% of the gap between high income countries and poorer countries, which I think is a big deal. It's not something that we... The education sector is rightly focused on improving teaching and improving access to school. But this really shows that there are these other factors which really affect how prepared children are to learn when they turn up in the classroom. It reminds me of when I was in graduate school, learning about deworming as being the big flashy thing, the most cost-effective thing you could do in education, which to most people working in education, in the Ministry of Education, was just bizarre. But I think there's deworming, there's lead, there's growing evidence about air pollution, there's all these other factors which affect children's readiness to learn.

So I think this is important. Perhaps eliminating lead altogether is a big ask. That's not something we can just wave a wand and easily accomplish. Unlike deworming, there's not just a simple pill you can take. We need to remove multiple sources of lead exposure. Here we benchmark both full reduction and some feasible demonstrated interventions and benchmark them against some education interventions. Two popular education approaches at the bottom, providing information about how the financial returns to more schooling or teaching at the right level, they're going to get you between a 0.1 and 0.2 standard deviation effect size increase in learning. So the complete reduction of blood lead would get you more than either of those interventions. But some of these other approaches, they're going to get you some gain. So soil remediation is the top there. That's quite expensive at a particular polluted site if you dig up and remove all of the soil. So perhaps not super cost-effective just on educational outcomes. But then some of these other interventions are not necessarily very expensive. So the leaded petrol ban, which has been accomplished, that had quite large effects on blood lead. Paint remediation, there's a lot of space left. There's lots of countries that don't have bands on leaded paint yet. So there's some really low cost regulatory

interventions there, which could produce these meaningful gains in learning, as well as all the health benefits.

So I'll conclude. I have some time, but that's cool. It's a big deal for education, and it's, I think, massively neglected and still huge... Most of, I think, the impact is potentially the health impact. There's a lot of uncertainty there. I think when people calculate big numbers for the ultimate impact of lead, probably most of the impact comes from cognition and productivity and IQ detriment. But there's also a big chunk, which comes from cardiovascular disease. We tried to put together the amount of international funding which goes on lead exposure, and it was, I think, about $15 million that we could find in recent years, which is just a tiny amount of money compared to the amount that's spent on health and education globally, including from aid, despite all the cuts. It's still a massively neglected issue. The next question, which we're working on now is, okay, so what do you do about it? What are the most important sources?

It's really tricky. We don't have a great understanding of what the most important sources of lead exposure are. There's, I guess, two big buckets, consumer products and then industrial sources. Even at that high level, it's not totally clear. But there is some real low-hanging fruit there. The fact that lead paint is not banned in every country. There's some great organizations working on this, but that's some real obvious basic regulatory issues, lead limits in water. There's a whole list of different regulations that exist in many OECD countries and exist in the US, but don't everywhere. National surveys have shown to be quite influential in Bangladesh and Georgia and Bhutan recently. I think the World Bank is supporting a survey in Indonesia. So in many countries, including the UK, where I come from, we haven't done a national blood lead survey in decades, if ever. And so this can be really powerful in showing policymakers like, wow, half of your children, a third of your children are poisoned, has shown to be really catalytic in creating action. And then, yeah, we hate to end every presentation saying we need more research, but we really do need more to try and understand those different resources.

And so we're doing work at CGD on battery recycling, understanding how much of a problem is coming from informal car battery recycling, which seems to be a lot. People just in their backyard are smashing open batteries which have lead in them in order to recycle them. And that dust just goes everywhere. And metallic cookware seems to be a really big one. So roughly half of the aluminum cooking pans that you can buy in Africa have an elevated lead level in them. Because people are making them with scrap metal, any metal they can get their hands on, that's a big constraint. And sometimes a little piece of lead will make its way into that stew. And so every time you cook with that pan, you get a little dose of lead. But we don't know how big of a problem that is. So lots more research. Thanks.

Deon Filmer: Thank you, Lee. And here, that's aluminum. Next up, we have Gabriela.

Gabriela Smarelli: Thanks. Thank you, everyone here and also online. I'm very excited to be part of this panel. Today, I'm going to talk about school violence, and I think everyone in this room will agree that violence is intrinsically bad. And evidence has also shown that violence has negative effects in child development, including in learning. In this presentation, I'm going to take you a step back to think about the measurement of violence. I'm going to be talking about how we can advance or improve the way we measure violence in schools, and I will be using evidence from Malawi. I will start by sharing with you three facts of what we know based on the existing data. First of all, we know that children are experiencing high levels of violence in school. If we take the case of Malawi, we observe that 42% of children have self-reported experience of school bullying. If we look at sexual violence, we observe that around 22% of girls have experienced sexual violence, and 16% of this is happening in schools. Another thing that we observe from the data is that most of this data is focusing on adolescents, so we know little about what is happening to the younger population.

A third key point is that the existing global surveys or international surveys that collect data on violence against children are using different methods to collect data and are doing this at different

locations. For example, existing surveys are done either using face-to-face surveys or self-administered questionnaires, and they are done either at the school setting or at home. But we don't have enough understanding on how this is affecting the estimates or the measures we have about violence. Across these surveys, and also among us that are collecting data on sensitive issues, there are some of the big challenges when we collect this data. One of the key challenges is that this might have some misreporting and will be collected with some degree of error. And there are different factors that influence this. One of them is that children that experience violence might not want to speak up or share the truth of their experiences of violence in a survey. This might be explained by fears of breaches of privacy, retaliation, stigma, or embarrassment. This can lead then to have some systematic misreporting in our data. Another factor that can lead to some error relates to the fact that especially when we work with young children, there can be some misunderstanding of the questions, as well as some lack of attention when we are collecting this data, and that this also can result in having some error in our data, an error that is likely to be more random.

So addressing this and trying to minimize this error is of crucial importance, both to have more accurate numbers of the dimension of the problem, but most importantly, to ensure that when we are measuring the effectiveness of interventions tackling violence or targeting violence, we ensure that we are measuring this accurately and not with bias. So this research is going to think a bit more about this and it's going to try to contribute to this. What we are going to do is that we're going to try to identify which data collection methods would give us more reliable measures of a school bullying, corporal punishment, and sexual violence. We are going to focus on a younger population of children, specifically on children aged 8 to 12. The way we're going to do this is that we design an individual level survey experiment with four treatment arms, where children are going to be randomly allocated to any of these four treatment arms. In this, we vary two main components. On the one side, we vary the data collection method, where we use either a face-to-face method or something that is called an audio-computed assisted interviewing method. For this, what you have to imagine is a child that will have a tablet and a head set of phones, and they will be hearing a recording that is asking them different questions of violence.

This method is at the end providing a bit more privacy relative to a face-to-face method. The other component that we'll be varying in this SRV experiment is the location. In other words, the location where we administered the survey. This will be either at the school setting or at the home of the respondent. We decided to use these different methods of data collection to see to what extent varying different levels of privacy will lead to different levels of reporting of violence. Something important is that for this study, what we use as the control group is the group of children that were allocated to respond to this survey using a face-to-face method at home. We decided this to be the control group, considering that a lot of surveys, including DHS, prioritized this type of data collection. Before showing you the results, two points that I wanted to highlight, include that we don't observe any attrition, differential attrition or across our treatment arms. Another key point that I wanted to highlight is that we took into account safeguarding measures when we were collecting this type of data. I'm happy to answer questions about this at the end.

What do we observe in the results? I'm going to be showing you two graphs with the key results. This first result is showing you measures for violence related to physical and emotional violence. Here what you're going to see is a percentage point difference relative to the control group that in this case are the children that responded to this survey using a face-to-face method at home. If we start by observing the first part of the graph, in that case, we are comparing the face-to-face survey at home relative to a face-to-face survey at school. So mainly here we are varying the location of the survey. And as you can see here, location doesn't seem to matter that much, particularly for cases related to reporting of corporal punishment or emotional violence from teachers. We do observe that at a school, we observe that it's eliciting higher responses in terms of more experiences of physical bullying. Now we're going to talk more about ACASI. If we compare the face-to-face at home relative to the ACASI method, this is the one where children are listening to the questions in

an audio, what we observe is that regardless of the location, there is higher reporting of violence in the ACASI method.

For example, if we look at the case of physical bullying, we observe that there is a nine percentage point increase when we are collecting the data with ACASI. This is approximately a 20% increase in the reports of physical bullying. Let me move now to sexual violence. In this case, it's important to mention that sexual violence considers different forms of sexual violence. It covers sexual comments, unwanted or forced touch, forced kiss, pornographic videos or nude pictures, as well as forced sex. First, we do the same. Let's compare face-to-face at home relative to face-to-face at school. Here we observe, again, that children seem to be more comfortable responding to the survey in the school environment, and we observe higher levels of reporting sexual violence in the school environment. We then compare the face-to-face at home with ACASI, and regardless of the location, we observe high levels reporting of sexual violence with a CASI. In fact, here, the disclosures of sexual violence are very high, more than 100%, and even four times higher in the case of sexual violence perpetrated by members of the school staff. So two key points to mention at this stage is that it seems that location does matter and that children are feeling more comfortable reporting these experiences in the school environment.

Some of you might be asking yourself, why the school versus the home? Something that we identify in this study is that in the school, there were more spaces to conduct the survey. For example, 50% of the numerators end up doing the survey in the schoolyard. At this point, it was very clear for the child that you were far away from the listening range of others. This provided, let's say, higher privacy than perhaps what could be achieved at home, where the survey could be either inside the home or at the front door. And it was perhaps harder to show that you were farther away from the listening range of others. Keeping that in mind, I will now briefly discuss the cost effectiveness of these different ways of collecting the data. And here what I'm thinking of effectiveness, what we consider to be more effective is the fact that we are able to elicit higher reporting of violence. And of course, here we are assuming that this higher reporting, it's coming from accurate reporting, meaning that we are closer to the truth of what are the levels of violence. And what we observe is that increasing by one percentage point the probability of reporting violence costs around $5 when we do this using an ACASI method in a school, and the amount almost doubles when we do the same, but using a face-to-face method.

So this is indicating that ACASI seems to be also a cost-effective tool to consider when we are thinking of how to collect data on violence. Before concluding, I would like to also discuss two brief points. The first one is that perhaps also some of you are thinking to what extent ACASI has actually given us reliable measures, do we worry that perhaps these are not accurate? Here we do several robustness checks. One first thing that we find is that under ACASI, there is a lower social disability bias. So this, in a way, is telling us that it's likely that with AC ASI, children feel more comfortable telling us their true experiences of violence or have a lower incentive to misreport. A second thing that we do observe also in ACASI is that more children fail attention checks that we put in the questions, and that there are more children misunderstanding some of the questions. We, however, do a series of tests to see to what extent this would affect the results I just explained, and we observe that even when accounting for this, our results hold. A second question that you might have is how do these numbers that I spoke about today compare to what we observe from existing surveys?

To think more about this, we look at the data from 15 countries that have a survey that is called the Violence Against Children Survey, and they have a question on forced sex. We took that data and we observed that based on that data, 1 in 500 children reported they experienced forced sex. If we assume the same levels of underreporting that we observe in this study, those numbers should actually be something closer to 1 in 50 children that have actually experienced four sex. It's possible that these numbers are still a conservative number because in our study, we focus on a younger

population of children, and forced sex tends to increase with age. So it's possible that these numbers are even conservative. So just to conclude, two main messages. The first one is that the location of the survey seems to be important. It is important to consider spaces that provide higher privacy. A second point is that ACASI seems to be effective in eliciting higher levels through reporting of violence, and it's also more cost-effective. This, however, should not be interpreted as saying that ACASI should be the gold standard, but instead that relative to face-to-face, ACASI might be giving us lower systematic misreporting or underreporting of violence, so it's worth considering these different alternatives when we are thinking of collecting data on violence against children. Thank you.

Deon Filmer: Okay, let me invite all the panelists up. Can I just have a show of hands who is interested in asking a question? I just want to get a sense of how many questions there are in the room. Okay. So I'll refrain from asking on my own then. Yeah, so we have some time, so we can take a couple of rounds. I think we'll do what previous sessions have done, which is to ask, maybe take three or four questions. I guess since we have four panelists, maybe four. Please keep it short, keep it as a question, if you don't mind. Okay, raise your hands again. And then we have one there. Although you've asked a couple of questions, I'll come back to you. I'll give somebody else the first one. So over here, and then we'll go over there, and then maybe we'll go there, and then one more this side over here. Okay.

Audience 1: Yeah, I'd like to ask Dev to comment on how this method compares to the HLO, the Harmonized Learning Outcomes approach, which has been used by the Bank and underlined Human Capital index, and was also mentioned by Lee in his presentation.

Audience 2: Thank you. I have a question for Gabriela. I'm [unintelligible] also from the DEC Development Impact team. I wanted to ask what you thought of practices within the self-reported surveys, like list experiments, changing the order of the questions, and other techniques you can use to reduce the social desirability bias, as opposed to using the computer techniques, and whether you have looked into it, and have you seen differences in self-reported surveys that use these techniques versus ones that do not?

Audience 3: Hi, I think my question is for Gabriela.

Deon Filmer: Sorry, could you stand up?

Audience 3: Oh, stand up. Sorry. Hi. My question is for Gabriela, and I'm going to be quick regarding violence, such an important topic, especially within our community, the African community. So I wanted to ask if I missed it. I am so sorry, but I wanted to check on the corporate punishment if it was only on teachers, but not family members. If it wasn't from family members' worst parents, do you have... is there an opportunity to extend it, to look into family members' violence that maybe could trigger the bullying that we see at school, maybe not for corporate punishment or sexual violence. And also to tap into Dev's presentation, do you think that the household income or maybe the geographical context also impact the level of violence that we see in school? Thank you.

Audience 4: Yes, I had a question for Dev, which is, just where do Bihar and Florida fall in the relative distributions of sales for the two countries? And if you'll permit me, organizers, to have the privilege of squeezing in a second for Christina. Just thinking about cognitive endurance, when I think about... So I grew up in Lucknow, actually. I spent seven years of my childhood in Lucknow, and went to a small private school there. When I think about some of my friends' lives, in particular, compare that with my seven-year-old son's life here, their life was much harder in many ways. Is that something we should be thinking of as amplifying or competing with the cognitive endurance you talked about?

Deon Filmer: Okay, maybe start with Gabriela and work back this way.

Gabriela Smarelli: Sounds good. Thank you for the two questions. So on the first question, there are definitely other methods that can be considered when we are trying to collect this type of data. For this specific study, we also considered whether list experiments would be an option. Considering that with list experiments, they are more difficult to explain and sometimes to understand, particularly for young children, we decided not to test these, but it's definitely something to consider. And overall, I would say that when thinking of which methods to use to collect this type of data, it really depends on the population you are working with. But there is also other good research, comparing these experiments face-to-face, particularly on the topic of domestic violence, and also showing some mixed results, but something worth to explore, thinking about that. On the other point, we only focus on corporal punishment from teachers. So this study is mainly focusing on forms of violence that happen in the school environment, but definitely something important to also study and also study as you explain these connections between what happens in the school environment and how that might result in changes in bullying in the school environment or differences in the likelihood of experiencing sexual abuse.

So another point that is important to study. And you had another question at the end... There are studies showing that these are drivers that explain likelihood of being a victim of bullying and also sexual violence. So there are already some studies making that connection as these being drivers that might contribute to more exposure of violence. Thank you.

Deon Filmer: Okay. Christina.

Christina Brown: Thank you so much for your question. This is definitely something that we had in mind when we started doing the experiment that the types of day-to-day lives that these children have, it's very distracting environments compared toa lot of similar students in primary school in high-income settings where you'd actually have time outside of school to focus for extended periods of time on homework or other assignments. These kids just weren't getting that type of practice when they're at home. And then in the classroom, they're getting a few minutes per lesson of a focused individualized practice. So it feels like the baseline level of time on focused practice is very low, and there's a tremendous amount of distraction that's preventing this extended focused practice. So I think the baseline level is low, which definitely contributes to lower overall cognitive endurance. And part of what we had in mind when we were designing the intervention was keeping in mind that their time is scarce. There's already a lot of other things on teachers' plates in schools, on students' time once they're out of school, or supporting their family or doing other things. And so we were hoping to basically say: "Can we take an existing class, an existing lesson, and tweak it in some way?" So that, for example, what the teacher is doing, they still get to teach the same content, but we can change something about the pedagogy so that they're also able to build this other skill. And I think that's something, hopefully, for schools to think about as well, too, is what are the types of teaching practices that allow you to not only increase literacy and numeracy skills, but also build these fundamental socioemotional scores. It doesn't have to be one or the other necessarily.

Dev Patel: Great. Thank you so much for those questions. So for those of you who don't know, the HLO is this very impressive data product that's released by the World Bank and covers a much larger set of countries than what Justin and I have covered in our paper. I think if you were just to directly compare, for example, the average test scores in our sample to those, you would see that in general, we seem to find that the low-income countries tend to be doing worse than it was estimated by the HLO. For example, I think Chad is scoring about 0.2 standard deviations lower in our measure than what the HLO is estimating. But the second thing I'll say is that one of the key advantages of our approach is that it really lends itself to looking at the microdata. We can really examine not just country averages or certain moments of this distribution, but actually looking at how private schools or the gender or these other things are really interacting with test scores. The third thing I'll say is that I think one thing that we emphasize in this talk is that there's a big learning crisis going on. I

think you didn't need me to tell you that. I think you all knew that beforehand. We didn't really need to have that test where it did show that. But there are some other policy implications, I think about our work. And in particular, one thing that I think was really salient to me from the meta-analysis that Lee was showing is that Abhijit Banerjee once famously said: "There's nothing standard about a standard deviation."

And there is, I think, a challenge that we have when we're comparing studies across places for all measuring test scores in different places or running our own interventions or things that... Christina does an amazing intervention, improving cognitive endurance in India, but we want to compare the test score effects to that, to a different intervention, to deworming, for example. We can't really... There's nothing sensible about just comparing the standard deviation units. And so we can actually use this type of methodology that Justin and I have talked about to try and link all these test scores on a common scale. So all you have to do is next time you go out and run your intervention, including a few questions from one of these tests. And suddenly we can actually just report all of our standard interventions on TIMSS units, and then we can actually have these on the same scale. And then I think that's related to your question, [unintelligible], as well. So one thing that we're limited by in the existing data of these sampling frames of these tests is that they're concentrated essentially among specific grades in a specific year in countries. And so when we went to TIMSS, for example, it was fourth graders in 2011. So we're going back to Florida and doing all middle school or a bunch of grades in Bihar. It's not quite obvious how to exactly compare those, but you do see a huge range and actually quite a bit of overlap. In general, it seems that Bihar, in particular, of course, at the lower end of the distribution, I think one big thing that emerges from Justin and my work is that there's a big empty spot on that map for India and China. Lack of standardized tests for those countries is a huge policy problem for all of us who care about kids learning.

Deon Filmer: Thanks. I'm going to go down and I had a few questions myself, so I'm going to take the chance to insert them. I actually had the same question as Gabriel for you, obviously coming from the World Bank. But I had another question, too, which is, you infer you didn't say that the lack of... Would you agree that the lack of association between national income and test scores is because of the quality of education service delivery across the countries? I mean, that's the obvious inference, but I don't know if that's true or not. And then the other question for you, Dev, is on the inequality result, could it just be that when the Gini is higher, the income differential is just higher between the top and the bottom, and so you get this coefficient? That's a question. To Christina, my question is, as you were presenting this, I was thinking to myself, wow, you're just undermining the Big Five interpretation of perseverance as a trait, and you're saying it's a skill. But then at the end, you treat perseverance as different than what you were doing. And then you also refer to sustained attention.

So I guess, would you mind just elaborate a little bit on how these are different or similar dimensions of the same or different things. To Lee, I know this is completely unfair, because if you could have done this, I know that you would have done this, which is the obvious cost-effectiveness question. And so what are your thoughts on cost effectiveness of abatement and how that would relate to your work? But maybe the fairer question is... You project this onto HLO numbers which are pegged either end of primary school or somewhere late in secondary school. And then that got me thinking about, well, what's the timing of exposure and what grades are these measures actually being measured? And is that a fair projection from the studies, which I don't know, but you do, to how you're trying to quantify the impact? Is that fair? And then to Gabriela, you're comparing a lot of survey methods. I guess I'm curious about, and you probably did this in the piloting, but what about non-survey methods and social workers? How do these surveys compare to what you get out of a non-survey type exchange with the respondent? I don't know, should we start on this end this time? Then keep your questions because we'll do another round in the room.

Dev Patel: Great. Justin and I are big fans of cross-country regression. We did do the regression of test scores on education spending per capita, and there's a positive relationship there. To whatever extent, I think I'll maybe leave the interpretation up to the rest of you, but we do find that positive correlation. Then on your second point, I think one of the key advantages of that regression that we're allowed to run is we can put everything in the same unit. So it's household income in PPP dollars, and it's test scores in TIMSS units. And so the interpretation of that coefficient in that Gini plot is for a certain log income increase, what is the increase in test scores. And so I think differences in the overall distribution won't really matter there. It's going to all be per the same dollar income.

Christina Brown: Great. Really great question. I'm not a psychologist, so I will share some of this, but the way I think of separating these different concepts. So in terms of the Big Five personality traits, I don't know how much evidence we have of the extent to which this can be developed through training otherwise, but in the category of executive functioning skills, so working memory, sustained attention, which I say is very similar to cognitive endurance. I think that psychologists measure sustained attention a little bit differently than how we're measuring cognitive endurance. But I think of them as very fundamentally very similar concepts. So I'm happy to have you use them in exchange. And then perseverance, I also put in the same category of other executive functioning skills that there is some evidence for working memory, perseverance, and now sustained attention or cognitive endurance, that this can be developed through practice. So it does seem like this is an area that's their scope for schools to potentially influence or parenting, for example. And in particular in our study, the way we try to separate this is that we have another variation that we introduce. So for these tests, we also have some where kids get a toy prize based on their performance.

And you see this really interesting pattern that when you have an incentive, kids work harder at the beginning of the test. They do better at the first few questions. So incentives work, yay, economics. But you actually see them decline even faster over the length of the test. They start to really flag very quickly. They're really motivated for the first few minutes, but then they do quite poorly at the end of the test, even when there's this incentive prize. And this doesn't crowd out any of the effects of our intervention. So it seems almost like this is a separate mechanism. There's the motivation channel coming from the incentives, and this is separate from the cognitive endurance or sustained attention muscle. And I think of it similar to, you could give me a million dollars to... I run a four-minute mile, I'm going to have a lot of motivation. It doesn't mean I have the musculature to be able to perform that. We think of that similarly with sustained attention, that you can have the incentives to want to complete some tasks, but you need the underlying attention muscle to be able to then capitalize that and focus over time. So that's how we think of separating them.

Lee Crawfurd: Thanks. The first question was about cost-effectiveness, and I think we looked for all the studies we could find on what works to address lead exposure, and there's hardly any, so it's difficult to draw firm conclusions. There are some fantastically cost-effective stories out there. So the story of the banning of lead in petrol and gasoline globally, the World Bank played a big role in that. It happened in the US and in other high-income countries. There was a big conference organized in Africa, and essentially, they all agreed: "Oh, yeah, we should do this with support from the UN and the World Bank", and they all just did it in the space of a few years. And that cost very little and had a huge, huge impact. Similarly, more recently, the NGO is working on lead paint regulation. It's not that there's a really big, powerful pro-lead paint lobby. They just don't realize that it's that big of a problem, and it costs a bit to reformulate the paint, or there might be one or two holdout companies, but it's not that big of a... It's an amazing rare occurrence where actually a bit of data and evidence really can swing policymakers because it's just they haven't thought about it and don't have a strong prior. And then in Bangladesh, where it costs next to nothing for one PhD student to dig around and try and find what the key source was.

And she discovered they were adding lead to tumouric. And so they just went down to the market and they confiscated it all. And the government didn't know about this, and was very happy to

enforce the existing regulations. And the tumouric producers were very happy to stop doing it because they didn't realize they were poisoning all of their customers' children. So there are these phenomenal opportunities out there, versus cleaning up decades of old mining site where there's just soil across the whole town as there is in Kabwe in Zambia. Currently, they're trying to sue Anglo-American and the London Court to try and get them to pay for that cleaner. But that's expensive. So the worst site, it's not always going to be super cost-effective, but a lot of the regulation probably is. And then at the age of the kids, I think this is lifelong impairment to your brain. And so we see effects on teenagers with crime and violence. I think these studies do show a range of ages of the children. I think we don't know exactly what the cutoff point is. We think it matters most under six for children, as your brain is still developing most rapidly. But we don't know if you remove lead exposure for a 10-year-old, is that going to help them? It probably will, but there's a bit of a lack of clarity there.

Gabriela Smarelli: Thanks. That's a great question. In our research, something that we did is that our team included not only enumerators, but social workers as well. This happened during data collection that every child had the opportunity to speak and was offered the opportunity to speak to a social worker. And that gave us quite unique information. And in total of the 6,000 children that were surveyed, 25 agreed to speak to a counselor. And that gave us unique information in the sense that through these one-on-one conversations where you could build more trust, children also opened up about other experiences of violence that happened outside the school, including domestic violence. We were also able to identify if children were misunderstanding some of the questions. So that's why at the end of the presentation, I mentioned that we also did some analysis for correcting for those children that misunderstood some of the questions. So these other alternative methods of trying to do more one-on-one conversations with social workers are definitely a good option to try to understand how to get this measurement right, and how to ask these questions in the best possible way to ensure that children are understanding. And this makes me remember the paper by Chris Blattman and others in Nigeria that I think they also did use a similar method to do this. So something that we can consider as well for this type of studies with children.

Deon Filmer: Thanks. Okay, so let's do one more round in the room. I know you had a question. Maybe three questions total, and then we'll do... Okay, two questions, I'm told. One here, and then maybe one on the side. Keep it short, keep it a question, and then we'll have a response.

Alberta Hagen: Thank you very much, Dean, for the opportunity. And thanks for the brilliant presentations. A little question to Dev. You indicated that girls outperform boys across all income levels. But then when it comes to maths, they only outperform the boys at the lower income level. I just wanted some explanation about this finding. Thank you.

Deon Filmer: Okay. Any questions on this side? Okay, over there.

Audience 5: I'm curious about the new wave of education, especially when it comes to AI, and basically you have to review test-taking as a whole, intrinsically. What do you think that is going... How is that going to affect standardized testing? And how are the results that we're seeing across different countries? Maybe they should be different because the skills that you need in places are different. Is that something that maybe we should take into consideration when comparing across countries?

Deon Filmer: I guess both of those were sent your way. So why don't you take a shot?

Dev Patel: Great. So this first question, it's a great question, and I don't have any good, real answers. Maybe I'll just leave you with one striking statistic that's stuck in my mind, which is, guess what country in the world has the largest test score gap between boys and girls for fourth graders? It's actually Saudi Arabia. And girls are way outperforming boys there. So girls are scoring about 0.4 standard deviations higher than boys in Saudi Arabia. And I've never actually quite understood

the origins of that statistic either. So I think there's lots of open questions here that would be great to answer. I'm sorry, maybe you could say your question one more time.

Deon Filmer: AI and then is it okay to have these differences because different things are needed in different countries?

Dev Patel: Yeah, it's a great question. And again, I don't think I have any really good answers to predicting the future, especially on the AI side. I think one thing that's really emerging from what we're seeing is that there's a huge amount of inequality that does seem to be somewhat systematic. And this is related to what you were saying earlier, in ways that we should be able to predict. And so we should be able to look at those different gradients in which countries are doing better and worse and try to understand a little bit better about what features of the education systems are worse. I know, for example, Jishnu has been doing some really amazing work on thinking about education problems at a systems level. I think, for me, that's actually certainly the most exciting type of work out here.

Deon Filmer: Okay, I'm going to weigh in a little bit. I'm not going to weigh in on AI. We had a whole session yesterday. On the other one, I mean, Lee showed the chart where 29% of kids, age 10, couldn't understand a basic sentence, right? They couldn't. So that's a pretty foundational skill and one that's required pretty much everywhere. So parts of this just apply everywhere, I would argue. Yes, there might be more subtle skills that vary across locations, but I don't think that's really what's driving a lot of what was here. Anyway, sorry, you didn't want to hear me pontificate. Please join me in thanking the panel. That was great. Thank you very much.

[END OF TRANSCRIPT]