Education - Evidence and Recommendations

External Comments

Vadim Albinsky, Applied Researcher

These comments were made to an older version of the report, and were invaluable in improving the current version. My responses are below each reviewers'.

Lant Pritchett:

This is the first of the EA/GiveWell type reviews of the evidence I have seen so it was interesting to see what is being done to reach conclusions about specific interventions.

Overall I find this persuasive relative to the other "preferred" interventions.

I have comments in three buckets: (a) additional papers you may want to see/cite integrate into the argument, (b) some paths to enhancing persuasiveness (which dovetails into a discussion of method), and (c) some puzzles remaining.

(a) Some papers (and points about them)

(i) The Pakistan RISE team have a new paper showing the long-term effects of cognitive skills and non-cognitive skills on labor market outcomes that was just accepted in the JDE (attached). While not directly "interventions" to outcomes, it is one of the few papers with long-term follow up from scores as children to adult outcomes.

(ii) A recent paper (still in process I think) by Noam Angrist and a co-author look at the huge variation in intent to treat estimates of impacts. I think this paper is generally relevant to the question of how to assess the data from experiments as they show order of magnitude differences in ITT estimates--but they suggest this is nearly all variation in implementation as "treatment on the treated" is quite similar.

(iii) One reason for my nervousness about reviews of evidence is that the "impact" or "impulse response" surface over the design space can be quite "rugged" with non-linearities and interaction effects. This Kerwin and Thornton paper about literacy programs in Uganda is a cautionary tale where a big impact--but at high cost--went to very small (even negative) impacts when the "same" program was done in a cost effective way.

(iv) As you might guess, I find the whole "methodological skepticism" approach dubious both as a way to do economics and, even more so, as a way to make pragmatic decisions. Attached is a recent paper of mine arguing that the observed heterogeneity of non-experimental "correlational" results is a fact the "best available understanding" must encompass. The "give weight to the 'best' studies" to "systematic reviews" has no coherent defense.

(b) persuasiveness

I obviously like your approach of saying "Hey, what makes sense overall from the weight of the data even when we don't have a single "killer" study."

In that context though I am puzzled by your more or less complete dismissal of the Mincer evidence (or, your intermittent use of it).

That is, your question is: "If an organization could reliably produce at some stage of a child's schooling an X standard deviation increase in measured cognitive skills at cost of $Y what is the increment X to the income of those individuals as adults, and are there demonstrated interventions of X, Y and Z that are are persuasively cost-effective as the EA/GiveWell preferred interventions?"

I guess how I would start making that case is that the Mincer regression (or some variant of it) is perhaps the best documented fact in all economics (only the Engel curve comes close) in the sense that people with more schooling make more money in a very wide variety of settings (lots and lots of countries, lots of countries with repeated estimates over decades) and the variance of those "increment to wages from a year of schooling" estimates across time and place is actually, to my mind, remarkably small, centered on 10 percent from an incremental year, with very few below 5 percent and very few above 15 percent.

One causal interpretation of that robust fact is that the labor market rewards cognitive skills, L, and a year of schooling, S, conveys those skills, If we assume that (a) the typical Mincer increment is 10 percent and that (b) the typical increment to L per year of S is about .33SD of measures of L and (c) that 100 percent of the wage return from S is mediated by L then the wage increment to L is about 30 percent per SD of L.

(a) and (b) are just facts (that vary a bit across countries and stuff) and hence only (c) can be doubted. There are two doubts: (a) the observed Mincer coefficient is due to selection (either ex ante (people who would have higher wages do more S and get more L) or ex post (employers use S as a signal of productivity) and not causal to a year of schooling and (b) that more happens in a year of schooling that just L and hence the causal impact of S on w is only gamma percent mediated by L.

All of this I guess dovetails into a method discussion and two ways I think the whole RCT movement has led the EA movement astray.

That is:

(i) judgments should be made on the basis of "understanding" and evidence enhances understanding. But understanding is, in one way or another, a "model" (even if not fully specified) about how the world (or the relevant bits of it for the judgment at hand work and are interconnected. The RCT movement (at times at least) acts as if we can bracket any "understanding" and just "rely on the evidence" which is, to my mind, beyond stupid. It is as if they want to roll the clock back from chemistry to alchemy (which was highly experimental, just based on a wrong theory). I think one cannot legitimately or persuasively say "I am basing my judgement on whether L increases w on the following three or four experimental studies" and not ask yourself "Is my judgment consistent with a viable interpretation of the facts about associations revealed from thousands of Mincer studies." And I don't feel one can be "neutral" or just "bracket" this by saying "Hey, I am just going to ignore those studies because there is a story in which the association "might" be due to this or that and not causal."

(ii) The second general method-like point is that I think the RCT movement has tended to encourage people to conflate "methodological" scepticism as a stance in a certain way of doing academic research (as, vaguely a kind of Wittengenstienian language game that is played according to some rules) versus actual "pragmatic" scepticism which is presumably about actually making real world decisions.

My blog on "feigning ignorance" discusses this in the context of a paper that showed with an RCT that if schools were closer to kids they would go to school more. They justified this as a "finding" worth publishing in a top economics journal because the literally thousands of studies showing this--and the economic model in which people do more of activities that are cheaper (all else equal) could be completely ignored because they could not, in and of themselves, resolve causality and one could tell a story in which the association was "reverse causality" that schools were built near kids who wanted to go to school. But never mind that (a) that story was a fantasy in the mind of someone concocting a possibility that justified a "methodological" scepticism and (b) that if one were constructing an actual Bayesian belief of whether kids would go to school more, and by how much, if schools were closer one would rely on all of the evidence and of an understanding of the theoretical causal mechanisms. So this paper had the flavor "We are providing rigorous evidence that the impact of distance to school is not zero so if that were your prior you have ample justification to change that prior." But no one had that prior in the real world, this was all an internal language game that academics were playing in which one was allowed to pretend that it was reasonable to adopt completely crazy beliefs as if they were your true beliefs.

c) All of that said, your particular argument is quite intriguing and much harder to make the case for that the easy case.

That is, the easy case would be ""If an organization could reliably produce at some stage of a child's schooling an X standard deviation increase in measured cognitive skills at cost of $Y what is the increment X to the income of those individuals as adults, on the presumption that increase persists from the intervention to the labor market and beyond--and are there demonstrated interventions of X, Y and Z that are are persuasively cost-effective as the EA/GiveWell preferred interventions?"

In this easy case you just have to make the case that (i) increases in cognitive skills increase wages, and (ii) the interventions raise cognitive skills into the labor market.

But you are making the case that *even when the fade out of the gains from the intervention are zero there are still labor market effects.*

This latter claim makes your job so much harder as, in some ways, it almost exactly reverses the case that would be made without fade-out. That is, if I am making the zero fade-out case then the PIACC studies are pretty persuasive that cognitive skills, of the type academic assessments assess have a labor market return. But with zero fade-out this is irrelevant as the d adult L/d Intervention could be zero and you still want to make the case that d adult w/ d intervention is positive.

I guess as a person who wants a "empirically validated theory" as a set of causal statements that encompass the evidence as a basis for decisions I would be much more persuaded about what the causal mechanisms might be that lead to the effect on wages even conditional on 100 percent fade-out of the gain to L from the intervention.

One final point, which is that the whole point of Kaffenberger and Pritchett was not particularly to demonstrate a learning to good outcomes impact but the deeper point that with S and L to wages there are stories of signaling and stuff that create some scepticism about causation but that with maternal education and child outcomes these stories are irrelevant as there is no "signaling" possibility. So in an indirect way, that we did not apparently make sufficiently clear, we wanted to say "Hey lets add to your Bayesian beliefs about how and why schooling adds to earnings by examining a set of phenomena in which stories that operate via signalling have no (or much less) plausibility."

That said, I was actually quite surprised by our empirical results. I was thinking that by bracketing out signaling we should get something more like 100 percent L and 0 percent S. I can see just sitting through a year of S has wage benefits for a variety of reasons via signaling or actually making you a better employee by having tolerance for doing pointless things that people in authority tell you do to but I have a hard time why, among women with just primary versus women with no schooling a year of education that did not enhance learning would make you more likely to have your child live.

I realize this was perhaps not the email you expected, but I think the case you make is pretty good, pushing back about just financing trivial things is important, but could be more persuasive by stating more clearly not just that "various sources of evidence are consistent with this" but being a bit more specific about the overall causal narrative of why high impact on skills interventions can have pretty substantial wage effects, which is pretty easy with low fade-out but, in my view, much harder with large fade-out and making it clear how this distinction affects who we read which source of evidence (e.g. spending on wages in the US or PIACC (which don't seem fadeout evidence) versus Head Start or Perry Pre-school or Star.

Response from Vadim to Lant Pritchett:

a) These papers have really helped fill in some context for me. Thanks so much for sending them.
i) I don't think you attached the Pakistan RISE paper on cognitive skill impacts. Would you mind resending it?
ii) Thank you for this. I've included a link to this paper in the report, and made an update to the "discount for small-scale studies section" to consider uptake and fidelity. Unfortunately, we usually only have limited information on this for most interventions, and the best we can usually do is to lump them into the discount we apply when the initial evidence is for smaller evaluations, and the final program is scaled up. We assume the scale-ups have lower fidelity and uptake, and they usually have more government employees (instead of nonprofit employees and volunteers). This paper will also be helpful for refining our cost effectiveness for TaRL Africa (one of our top education recommendations).
iii) We've definitely run into this phenomenon before, where smaller programs move to small or negative effects when done at scale. I think this is a big part of the reason we should apply such steep bayesian discounts when we extrapolate from evaluations of small-scale programs. This paper is a helpful reminder to require much higher effect sizes when we fund scale-ups of such programs, and to fund evaluations of their effectiveness after they scale.
iv) I think we share the belief that in areas where OLS estimates provide more studies, especially ones that are more geographically or otherwise relevant than RCTs, we should consider them in our analysis. This paper moves me further in that direction, and I will spend more time looking for OLS evidence in my future investigations.

b) I agree that the Mincer evidence is pretty compelling, but I choose to refer to it only as a supporting argument in footnote 3. Instead, I focus on Mincer-style regressions on PIACC scores (rather than years of schooling). A major reason for this is that the most cost-effective interventions we have come across focus on increasing learning gains, making the evidence around years of schooling less relevant (I've updated the report to mention this). A second reason is that, as you have pointed out in your work, schooling does not always lead to learning. As a result, the quasi-experimental and the cross-country GDP evidence about the impacts of years of schooling appears far less compelling than that for learning gains.

i) I think we agree on the correct methodological approach to the Mincer evidence. It certainly factors in as supporting evidence, and is consistent with a model of learning gains leading to higher income. We might just disagree on whether the traditional Mincer evidence or the PIACC Mincer-style evidence provides more support for my argument?

ii) Thanks so much for that blog post, I really enjoyed it and agreed with almost all of it. I will spend more time reviewing OLS evidence as a result of it in the future. I think to defend the EA movement on some (but far from all) of their methodological skepticism, I would say that they are often operating in areas where the bias in OLS results would include zero effect after a bias-bounding exercise. A secondary point is that even if the Afghanistan School Building paper wasn't a significant contribution to the literature, EA organizations sometimes fund RCT's like this of a specific nonprofit operating in a specific area they would like to fund, which I think you would agree should provide a significant update to cost-effectiveness estimates?

c) This section was very helpful in helping me clarify my thinking around fade-out. After more thought, I agree with your analysis that much of the evidence in this report (including PIACC) provides a counterpoint, and suggests that fade-out is likely limited. I think I over-relied on a couple of preschool studies, and over-extrapolated to longer-term fade-out from reviews of the evidence based on the initial fade-out after a year or two. I also think that my wording about fade-out oversold even my initial beliefs about the likelihood of fade-out to zero being a reality. I think this is a fairly counter-intuitive claim, and there is not nearly enough evidence to support it given low priors. I've updated the report to reflect what I now think about the evidence. There is likely some fadeout, but only a smaller chance that fade-out is complete. Therefore in our accounting for fade-out we should assume that effect fade-out to 50% over 2 years, but persist thereafter (as a sort of weighted average path we should expect based on the evidence).

I was just as surprised as you by the contribution of S in addition to L in Kaffenberger and Pritchett. I didn't focus on the relevance of this paper to signaling because I didn't assign signaling much plausibility in the contexts where our charities operate and was trying to keep the report from getting too long, but I agree that this is an important finding. I was even more surprised by the size of the effects on the mortality of beneficiaries' children, and that was what I initially thought might be most relevant to the investigation.

David Roodman:

I think it is very clear and thoughtfully done. I especially appreciated the creativity in the second half, where you combine the various strands of evidence and discounted them in various ways. And I learned about at least one study (in Colombia) I should have known about!

What I find myself thinking about is how OP has at present reached a different decision, however provisionally. And why. I think it's not about right and wrong ways of doing things; we all recognize the deep uncertainty. It seems like it's about what priors are doing the work.

Here's a curmudgeonly take on the evidence you present. I don't fully endorse it, but it may help you think about the role of priors. The Chetty and the Project STAR stuff has come in for an unusual amount of criticism, so maybe it's all mining (though I have never looked at it and have no idea!). Jackson et al. 2015 might have the same problem as Duflo (2001) (again I have little clear idea!). And anyway that's all in the U.S. The results from the lottery in Colombia look real (and new to me), though I think I read that most of the benefit was at the top of the income spectrum. And the Chile study I pretty much believe, as I have written. Same for the Ghana study, though that one emphasizes the zero-sum of the gains in the Ghanain context (which I think is worth bearing in mind, but shouldn't become a cudgel, because maybe that's true in a lot of contexts in the short run and less so in the long run). There's some small-sample, randomized Head Start stuff, and then a non-randomized long-term study that again may have the same problem as Duflo (2001), while long-term studies finding that the results fade out are not given as much space. Then there's all the international and cross-country education-and-income-go-together stuff—meh. So overall there's something here, but it's not super-strong. Plus, the two studies we looked most closely at (Duflo and Khanna) became substantially less convincing after scrutiny, especially Khanna, so what does that say about the rest?

I don't know if that skeptical shoulder shrugging represents the thinking of [OP], or even me for that matter if I were more deliberative. But maybe it will be useful for you in providing a contrast to your own synthesis.

I hope that helps. Thank you for your work.

A version of this report with David Roodman's specific comments and my responses can be found here. I have responded to why I do not think the criticisms that have been directed at the Chetty and Project Star papers should not significantly update our opinion on the specific results we are including in this report.