

Response to Blattman's Post on Why What Works Is The Wrong Question

Doug

2016-08-4

Last week, Chris Blattman published a long blog post titled “Why ‘what works?’ is the wrong question: evaluating ideas not programs.” In the blog post, which was adapted from a talk he gave at DFID, Blattman argues that a) impact evaluations should focus on deeper, theory-driven questions rather than just whether a program works or not and b) researchers should design impact evaluations to allow for generalizability by paying attention to context and running multiple evaluations in multiple contexts.

There's a lot to like in this post, but ultimately, it left me frustrated— not because I didn't agree with the substance of the arguments, but because I think he squandered a great opportunity to push DFID in the direction of better evaluation.

First, the bit I liked. Blattman's argument that researchers should design impact evaluations with generalizability in mind struck a chord with me. As Eva Vivalt and others have shown, context matters a lot for impact and extrapolating results from one context to another is really, really hard. Thus, if your goal is the creation of general knowledge you should care just as much about external validity as you should about internal validity: an extremely rigorous result in one context is of no use if policymakers can't figure out how the impact would like change if adapted to another context. As Blattman points out, this implies a shift in how we go about doing impact evaluations that seek to create general knowledge. Rather than one-off evaluations in one context, we should first think hard about context and then attempt to run multiple evaluations in multiple contexts all testing the same basic idea.

Yet while I liked Blattman's call for more attention to generalizability, I was disappointed that he didn't more explicitly tell DFID how to make this happen. Let me take a step back here and review how donors like USAID and DFID decide which programs to conduct an impact evaluation on (based on my, admittedly limited, experience). First, they design a program. Then, if the person in charge of the program is a fan of impact evaluations, that person will set aside a portion of the budget to run an impact evaluation. (There are exceptions to this rule, like DIV, but in general this is how it works.)

In this context, calling for donors to conduct multiple trials in multiple places of the same idea is a bit of a pipe dream: an evaluator would have to have to simultaneously convince multiple program managers not only to participate in an impact evaluation but also to subjugate project design and scheduling considerations to the needs of the impact evaluation. In my view, the way around this is for donors to clearly distinguish between evaluations whose main purpose is to improve the program being evaluated and evaluations whose main purpose is to create general knowledge. (For more on that distinction, see [here](#).) For the former, the existing method of leaving the evaluation decision up to program manager is, for the most part, fine. In contrast, for the latter, donors should first identify the big questions that they want to answer and then identify which programs can help them answer these questions.

Suggesting that an large bureaucracy should add yet another centralized, bureaucratic process is always a dangerous proposition but, in this case, I think it is necessary. Currently, at donors like USAID and DFID impact evaluations tend to be conducted on programs run by people who are sympathetic to impact evaluations rather than on programs for which there is little evidence or which should be a high priority for an evaluation for other reasons. This not only means that the type of ambitious, multi-site trials that Blattman suggests are infeasible, but that most impact evaluations tend to be up-or-down assessments of large, complex

programs which yield little in the way of useful results. A more centralized system for identifying priority research questions to be address by KFEs would make these evaluations much more useful.

Lastly, I was going to make one last point about Blattman's call for more theory-driven evaluations, but this post is already getting way too long so will save that for another post!