

October 2015

When Will We Ever Learn ... To Change Policy: Current state of impact evaluation

Yaojiang Shi

Center for Experiments in the Economics of Education
Shaanxi Normal University

Linxiu Zhang

Center for Chinese Agricultural Policy
Institute for Geographical Sciences and Natural Resources Research
Chinese Academy of Sciences

Scott Rozelle

Rural Education Action Project
Freeman Spogli Institute, Stanford University

Acknowledgements:

Thank you to the support of the Hewlett Foundation for the support for this work. The input of the 3ie and Hewlett Foundation groups that put on the workshop on “Looking Back at When Will We Ever Learn: a decade after” is gratefully acknowledged. The comments and advice of Annette Brown, Prashant Loyalka, Kristen Stelljes, Sean Sylvia and Howard White were particularly useful. Thank you!

When Will We Ever Learn ... To Change Policy: Current state of impact evaluation

1. Introduction

Nearly nine years following the release of the Center for Global Development's *When Will We Ever Learn? Improving Lives Through Impact Evaluation* report, and almost a decade into increased focus on evaluation among global donors, many in the research community are reflecting on the state of the impact evaluation field, whether the development community is learning what was hoped to from impact evaluations and where the future of impact evaluation leads.

This is also a time that 3ie is experiencing a transition between its first and second generation directors. As part of this transition, there is a natural tendency to reflect back on the experience of the last eight years and think about the years under 3ie's first director, Howard White. Hence, in this way, this essay is both a reflection of what is happening in the academic and policy worlds regarding the role of impact evaluation as well as a recognition of the achievements (and continue challenges) that can at least in part be attributable to Howard and his team.

As part of this reflection, in this paper we will explore the recent past, current status and future of impact evaluation of development interventions. We do not intend to focus on the (often polarizing) debates around methodologies. Instead, we propose to focus on how we can learn more to improve policies and programs with impact evaluation. We will describe the approaches that have been found to be particularly useful in not only carrying out impact evaluations, but, also in mentoring new groups of evaluators in developing countries who are able to take results, and in some circumstances, are able to turn them into policy change. Our goal is to inform the field of impact evaluation (researchers and other practitioners) as well as to look closely at what we and other impact evaluation groups are doing now and how we all can conduct even higher quality research and affect policy even more in the future. We also try to allow the reader to hear the voices of implementers (aid agencies, international organizations and policymakers) and understand how they view the recent past and continuing challenges of funding, designing and using IE.

In short, this paper aims to sum up successes and remaining challenges in the field of impact evaluation in the past decade so as to provide guidance for more effective evaluation and research-driven policymaking in the future.

To meet this goal and to understand the *current* state of impact evaluation, in this paper we both review the existing literature and conduct key informant interviews to provide an overview of the field of impact evaluation of development interventions, both in terms of conducting impact evaluation studies and in terms of using IE for policy change. We do so in two sections. First, we examine the gaps in our knowledge that still exist. We also examine the key elements that define any increase in IE "quality." This section of part one of the paper is essentially a desk review of the literature and an assessment of shifts in quantity, quality and the nature of IE. There will be three subsections: one review for health; one review for education; and one review for social policy. In the second section of the paper, we examine whether IE has not only emerged as a set of works that is teaching us more about what works, but also whether this newfound knowledge is being translated into policy. To do this, we report on the results of two surveys: one survey of Evaluators and one survey of Implementers. We also examine the challenges for both Evaluators and Implementers in combining IE research with efforts to change policy.

2. What have we been learning and what gaps remain?

In this section we examine the record of IE before and after *When Will We Ever Learn*. In particular, we examine the trends for IEs in the fields of Health, Education and Social Policy. We also seek to describe changes in the quality of IE and any remaining gaps in our knowledge.

2.1. IE in Health and Health Delivery

Possibly in no other sector has the growth of impact evaluation over the past decade been as large as it has been in health. In both quantitative and qualitative terms, the evidence base for social programs targeting health and health behavior has grown dramatically. Only 272 studies published between 1995 and 2004 meeting the inclusion standards of 3ie's impact evaluation repository were classified as "Health, Nutrition and Population."¹ As of 2014, the database includes a total of 1599 impact evaluations in this category, averaging 130 studies per year since 2005 (Figure 1).

In absolute terms, the increase in the number of quality studies (as identified by 3ie) has been largest in studies classified as "Nutrition" (1296 studies post-2004, compared to 301 until 2004), followed by "Preventative Health and Health Behavior" (998 studies post-2004 vs. 235 studies pre-2004) and "Health Services" (804 studies post-2004 vs. 191 studies pre-2004) (see Table 1).² In relative terms, studies classified as "Health Financing, Insurance and User Fees" and "Health Sector Reform" saw the most growth, with the number of post-2004 studies being 6.5 and 5 times what they were pre-2004 (albeit, these started from a low base and the cumulative number of studies remains small compared to other categories).

Methodologically, studies in health – both pre- and post-2004 – have been dominated by randomized controlled trials (RCTs). This likely results from both the longer tradition of RCTs in the health sciences and the fact that certain types of health interventions often lend themselves readily to randomization. The proportion of studies using RCT designs, however, has decreased as other methods have come into use (though alternate evaluation methods still only make up approximately 17.4% of studies in the repository as of 2014 – Figure 1).³ The proportion of studies using an RCT design is lower in all subject areas compared to pre-2004, apart from "Health Financing and User Fees" (Table 2). The growth of studies in areas such as "Mortality" and "Primary Healthcare" – where RCTs are often infeasible – seems to have been partly facilitated by the emergence of quasi-experimental methods.

Despite this significant growth in the evidence base, there are, of course, gaps that remain – both in methodology and focus. For example, researchers have drawn attention to the need for a more systematic and standardized approach to cost effectiveness analysis (Dhaliwal et al., 2011, McEwan

¹ In 2012-2014, 3ie commissioned a systematic search with the goal of collecting all impact evaluations ever published (in journals, books, reports or as a working paper) using "rigorous" methods. This yielded an initial repository of 2,400 impact evaluations meeting inclusion criteria. Inclusion criteria include 1) study published in English 2) the evaluation took place in a developing country and 3) the study must have used at least one of the following evaluation designs: a) randomized trial, b) regression discontinuity, c) propensity score matching, d) instrumental variables, or e) difference-in-difference. Randomized trials that only address biomedical efficacy of a drug or treatment were also excluded. See website (<http://www.3ieimpact.org/en/evidence/impact-evaluations/impact-evaluation-repository/>) for a full list of criteria. The repository is continuously updated with new studies.

² Studies are typically classified in multiple subject areas.

³ Some studies are classified in multiple method categories. The figure of 17.4% for the proportion of studies using quasi-experimental methods does not account for any studies that use these methods in addition to randomized evaluation.

2012: Evans and Popova, 2014). As the evidence base grows, it is increasingly difficult for policymakers to choose between competing effective interventions without a notion of costs (McEwan 2012). A recent review of the use of vouchers in health, for instance, noted a need to collect cost information for specific health goods and services under a variety of health financing models (Meyer et al., 2011). Another review of interventions to strengthen national health service delivery similarly noted a lack of reporting on cost effectiveness (Willey et al., 2013). This need for more attention to costs is not specific to impact evaluations in health, however; and indeed, there is growing attention to cost-effectiveness analysis largely arising out of the health policy literature (Jamison, 2015).

In terms of topical focus, a majority of evaluations in health focus on prevention, control and treatment of specific diseases or on the effects of specific health technologies while less research focuses on the health systems that would support these interventions at scale.⁴ There are critical evidence gaps, for example, regarding front line health workers (Frymus et al., 2013) as well as managers in health systems (Rockers and Bärnighausen, 2013). An ongoing, wide-ranging systematic review of strategies to improve healthcare provider performance rated only 14% of 490 included studies as having a “low” risk of bias (Rowe et al., 2014). This relative lack of quality impact evaluations in health systems research may be due in part to a lack of funding as donors in health prefer to support other types of health research that is seen as more prestigious or science-driven (Hafner and Shiffman, 2013). Health systems are also complex, involving multiple elements that affect ultimate service delivery (Hafner and Shiffman, 2013).

This relative lack of rigorous evaluations of approaches to strengthen health service delivery in developing countries is not for a lack of policy interest or resources devoted to this issue (Hafner and Shiffman, 2013, Willey et al., 2013).⁵ Strengthening health systems is one of the top priorities of the WHO and numerous multi- and bi-lateral organizations have been working in partnership with ministries of health in efforts to improve health service delivery (Hafner and Shiffman, 2013, Willey et al., 2013). These organizations even include those focused on specific diseases such as the Global Alliance for Vaccines and Immunization (GAVI) and the Global Fund to Fight AIDS, Tuberculosis and Malaria (Willey et al., 2013).

Learning from existing efforts to strengthen health delivery, particularly those at scale, requires collaboration between researchers and implementing agencies. With some notable exceptions, however, this is still rare. A recent meta-analysis of RCTs in schools found that 85.7% of treatment arms in RCTs of school-based health and nutrition interventions were implemented by researchers, while only 10.7%

⁴ The same idea was recently raised by Ruth Levine during an interview: “One of the central questions that many people are concerned about these days – and that came out in virtually all of the discussions during the meeting – is how to deal with the fact that donor funding tends to be earmarked for prevention, control or treatment of a particular disease, but those disease-specific programs don’t always support or strengthen the core health-related functions of government, and may even undermine those functions if separate information, personnel, supply chain and other systems are set up. This is the now very active debate about “vertical programs” and “health systems.”

⁵ “Interventions to strengthen national health service delivery” review (Willey et al., 2013): “Considerable in-country effort is ongoing to improve health service delivery in LMICs, often occurring in partnership between ministries of health and numerous multi-, bi-lateral and funding organisations, including but not limited to The World Bank, USAID, The Global Alliance for Vaccines and Immunization (GAVI), The Global Fund to Fight AIDS, Tuberculosis and Malaria, The UK Department for International Development (DfID), and Management Services for Health (MSH) (Vujicic et al., 2012). This review highlights that the accessible evaluations to address questions of effectiveness of approaches to strengthen health service delivery implemented at scale in LMICs using robust designs are considerably fewer.”

were government implemented (McEwan, 2014).⁶ A recent review of the effects of universal health coverage (UHC) schemes by researchers at the World Bank similarly noted that:

“...in terms of future UHC research, the review shows that most of the studies fail to involve evaluators from the start, which has led to weak evaluation designs to assess the impact of UHC schemes. Because of this, most evaluations are retrospective and do not use monitoring data. A better understanding of the effects of UHC schemes on financial protection is also needed to address several drawbacks of the current available evidence, and more and better evidence on the impact of health status is also required. A key step to address these difficulties would be to incorporate the evaluation in the early stages of the program and, ideally, simultaneously design the intervention and the evaluation. This would result in better and more meaningful evaluations, which in turn should contribute to enhanced UHC interventions” (Gideon et al., 2013).

2.2. *IE in Education*

The number of “quality” impact evaluation studies in the field of education (in developing countries) has increased markedly over the last ten years. Whereas only 92 studies met quality standards for inclusion in 3ie’s impact evaluation repository from 1995 to 2004, 512 studies met the quality standards from 2005-2014 (Figure 2). The last decade thus saw more than a fivefold increase in the number of quality impact evaluation studies in education.⁷

Not only has there been substantial growth in the number of “quality” impact evaluation studies in education overall, there has also been substantial growth within different areas of education. In absolute terms, the largest increases were in studies classified as “Pre-primary and Primary Education” (45 studies in 3ie’s impact evaluation repository from 1995-2004 versus 203 studies from 2005-2014), followed by “Secondary Education” (20 studies from 1995-2004 versus 99 studies from 2005-2014), “Vocational/Technical Education and Training” (6 studies from 1995-2004 versus 74 studies from 2005-2014) and “Educational Inputs” (12 studies from 1995-2004 versus 73 studies from 2005-2014—see Table 3).⁸ In terms of relative growth, studies classified as “Tertiary Education” (24 times), “Vocational/Technical Education and Training” (11.3 times) and “Non-Formal Education” (7.3 times) saw the most growth. It is important to keep in mind, however that the numbers of studies in Tertiary Education (1 study from 1995-2004 compared to 25 studies from 2005-2014) and Non-Formal Education (3 studies from 1995-2004 compared to 25 studies from 2005-2014) started from very low bases and that the cumulative numbers of studies in these two classifications remain small.

Methodologically, the growth in the number of impact evaluation studies in education has reflected a consistent growth in the use of both experimental and quasi-experimental research designs. From 1995-2004 as well as from 2005-2014, roughly three-fifths of the impact evaluation studies (those included in 3ie’s impact evaluation repository) were based on randomized experiments. By contrast, in both periods, about two-fifths of the impact evaluation studies were based (exclusively) on quasi-experiments. Within quasi-experimental studies, however, there has been a proportional increase in the use of regression

⁶ The remaining 3.7% were implemented by NGOs. For comparison, the same study found that instructional interventions and incentive-based interventions in schools were implemented by researchers only 22.6% and 23.5% of the time and governments 15.1 and 27.5% of the time.

⁷ Note, however, that the number of studies in education is markedly lower than in the field of health—between 2005 and 2014 there were more than 1500 in health and only 500 or so in education.

⁸ Studies can be classified in multiple subject areas.

discontinuity designs in the last ten years (2% of quasi-experimental study from 1995-2004 to 18% of quasi-experimental studies from 2005-2014) and a proportional decrease in the use of instrumental variables designs (35% of quasi-experimental study from 1995-2004 to 24% of quasi-experimental studies from 2005-2014).

Several authors have attempted to glean general lessons from the accumulated (experimental and/or quasi-experimental) evidence thus far. First, in their review of impact evaluation studies in developing countries from 1990 to 2010, Glewwe et al. (2013) suggest that school inputs fail to improve student outcomes while interventions that affect actions taken inside schools (e.g. interventions addressing school management or teacher incentives) improve student outcomes.

Second, according to his meta-analysis of educational interventions in grades 1-8, McEwan (2014) finds that computers and instructional technology, composite interventions that include some form of teacher training, contract or volunteer teachers, smaller classes or ability grouping, and teacher incentives have the largest positive impacts on student learning. By contrast, deworming, monetary grants, nutritional treatments, information dissemination, and school management have small or null impacts.

Finally, Murnane and Ganimian's (2014) narrative review argues that (a) reducing the costs of going to school increases educational attainment but not necessarily student achievement; (b) school inputs do not improve student learning unless they directly affect children's everyday experiences at school; (c) teacher incentives are effective when they are well-designed and when students have low levels of achievement.

Although the existing evidence reveals some general trends about what works in education in developing countries, much more evidence is needed to determine whether the impacts of most interventions are generalizable or whether (and how) interventions have to be adapted for specific populations and contexts (Glewwe et al., 2013; Murnane and Ganimian, 2014; McEwan, 2014).⁹ The vast majority of systematic reviews of educational interventions in developing countries suggest that many more studies have to be conducted before we can understand "what works" (in general) and "for whom" (specifically). This is true, for example, in research concerning vocational education (Tripney et al., 2013), the elimination of school user fees (Morgan et al., 2014), school voucher programs (Morgan et al., 2013), contract teachers (Kingdon et al., 2013), and so on. Furthermore, whereas the majority of impact evaluation studies focus on primary schooling, much less is known about which types of interventions are most effective for improving learning in secondary schooling and higher education (J-PAL, 2013; Clifford et al., 2013). Establishing a credible evidence base for secondary schooling and higher education has become increasingly important as developing countries are shifting attention towards expanding secondary school and higher education enrollments.

One challenge in widening the evidence base in education will be to establish more consistent standards for reporting. Fewer than half of the experimental studies (on improving learning in primary schooling) reviewed by McEwan (2014) report that a power analysis guides their choice of sample size. It is not

⁹ As seen in the Education Evidence Gap map on 3ie's website (<http://gapmaps.3ieimpact.org/evidence-maps/education-evidence-gap-map>), there are other areas, where it might be possible to know what works and what does not. Conditional Cash Transfers seem to have worked (at least in most countries) in improving enrollment and reducing dropout (although there are still exceptions, for example, a paper that is forthcoming in the Journal of Development Economics in the case of matriculation to high school in China). Our JDE paper shows they don't work in China). There seems to be consensus also that interventions to improve teacher attendance also help raise student attendance and reduce dropout.

always clear, therefore, if the lack of significant findings from a study is due to the ineffectiveness of the intervention or due to a lack of statistical power. In addition, up to one-third of the studies reviewed by McEwan (2014) fail to present data on sample attrition. Studies that fail to present data on (and appropriately account for) sample attrition may report biased estimates.

In addition to establishing more consistent standards for reporting, more attention needs to be paid on making the results of impact evaluation studies useful for policymakers. Even when conducting experiments within a particular country or region, researchers frequently rely on non-random, convenience sampling (McEwan, 2014). Researchers are thus unable to show how the results of their study generalize to the population that policymakers care about. In addition, when evaluating the impact of particular interventions, few researchers work with policymakers to bring those interventions to scale (McEwan, 2014). Most impact evaluation studies in education further only measure short-run effects (from the start to end of the academic year) and fail to measure longer run effects of intervention that may be of greater interest to policymakers (Murnane and Ganimian, 2014). As in health, impact evaluation studies in education provide limited evidence on cost-effectiveness (McEwan, 2012). In education, there are particular challenges not only with comparing incremental costs across interventions (as with health—McEwan, 2012), but also with comparing achievement gains that are based on different assessment instruments (and that are measured for different subjects and grades).

2.3. IE in Social Policy

Studies on social or welfare programs were included and excluded from our review according to the following criteria. First, we follow the general definition of social / welfare programs proposed by Toikka et al., (2004). As such, we only include studies with the goal of improving the livelihoods of the poor or vulnerable populations. Therefore, we only include studies that were conducted in developing countries. We also restrict ourselves to the theme areas that aim to increase income or bring about better economic opportunities for the poor. These theme areas include cash transfer programs, credit or microfinance programs, programs that target employment and labor market opportunities, social security or safety net programs, insurance programs and pension programs.

Second, when determining the studies to be included, we used several related keywords to do the search in each theme area using the 3ie database. For example, to search for studies on microfinance, we used micro-credit, micro-saving and micro-leasing (with and without hyphen) and microfinance. We also carefully avoided double counting the studies that appear in more than one of the keyword searches. We then conducted scanning to make sure that each study that came out of the search result did in fact fall under the defined theme areas.

In the past 10 years, there has been a large increase in the quantity and quality of impact evaluations (IE) of social or welfare programs. Based on the 3ie database, there were only 84 IE studies on social or welfare programs conducted from 1995-2004 (Table 4, row 7, column 1). The number increased by almost seven times by the following decade, 2005-2014 (row 7, column 2). During the 1995-2004 decade there were 8.4 studies per year; during the 2005-2014 decade the number studies rose to 64.9 studies per year.

Among the theme areas, studies on microfinance have increased the most. Studies on microfinance have increased from 10 studies during the 1995-2004 period to 148 studies during the 2005-2014 period (row 1, columns 1 & 2). The lowest rate of growth is found on studies on pension programs (row 6, columns 1 & 2). Only two impact evaluations were conducted during 1995 and 2004 and only four were conducted between 2005 and 2014.

The quality of the evidence has improved as well. First, more rigorous methods have been adopted to conduct impact evaluation. Many more impact evaluations were rolled out as full-scale RCTs. The total number of RCTs conducted increased from 24 during the 1995-2004 period to 276 during the 2005-2014 period. Microfinance is the area that has seen the most rapid growth. The number of RCTs on microfinance grew from only 1 RCT-based study during the 1995-2004 period to 65 studies after 2004. In other areas, there were fewer RCTs. For example, there were no published studies that used an RCT approach to evaluate a social security program between 1995 and 2004. Between 2005 and 2014, there was a small increase in the number of RCTs on social security programs (three studies were conducted after 2004). There have not been any RCTs conducted to evaluate pension programs either pre- or post-2004.

Quality has improved in other ways. For example, researchers have begun to adopt a wider range of analytical approaches aimed at enriching the insights that impact evaluations are able to generate. A few recent studies have employed mixed methods in impact evaluation (Kim et al., 2010; Miller, 2011). The mixed methods research combines quantitative analysis with qualitative research to provide in-depth and contextualized insights about program effectiveness.

Although still small in number, a few studies have also investigated the causal chain or mechanisms to understand why a program succeeds or fails (Akresh et al., 2012; Barr et al., 2012; Swain, 2012).

The increasing number of high quality impact evaluations of social or welfare programs has generated important evidence that was really unknown before the rise of large numbers of impact evaluations. Systematic reviews of the rising number of studies have shown that training programs and labor market programs are more effective on females than on males (Bouillon and Tejerina, 2007). Conditional cash transfer programs have been found to be effective in reducing the time the children have spent doing labor and have also led to increased household consumption (Kabeer et al., 2012). The conditionalities that are attached to the cash transfer programs have also been shown to be important in changing behavior (Hagen-Zanker et al., 2011).

In other types of studies researchers are also learning new things. For example, many types of insurance programs are now known to work to reduce poverty. This is true in the cases of life, funeral, death, crop and weather insurances (Radermacher et al., 2009). Micro-credit also has been successful across wide-ranging programs in promoting economic opportunities and financial outcomes for certain populations. However, the rising number of impact evaluations shows that this is primarily true for families with individuals that have received better education (Stewart et al., 2012). In addition, there is evidence suggesting that microcredit might actually be harmful to the poor.

Although much has been learnt from the existing impact evaluations, researchers have proposed several directions for improvement. Above all, it is thought by some that more studies on social policies are needed (White, 2014). Compared with impact evaluations for health and education interventions, the number of social or welfare programs that are rigorously evaluated is still small (despite the sharp increase in the past decade). In theme areas such as pension and social security, there has not been much effort to conduct impact evaluations. In areas that have attracted more research effort, such as micro-credit, still more research is needed to understand better how and in what circumstances the social programs can work for the poor (Stewart et al., 2012).

In addition, many researchers believe that more attention needs to be paid to non-compliance issues in impact evaluations (White, 2014). The nature of non-compliance needs to be understood to answer the

question of whether the program has targeted the intended population (Hagen-Zanker et al., 2011). There are whole sets of causal chain analyses that need to be done to understand why it is that only a small share of those eligible for programs actually benefit. In order to inform policies, it is also important to conduct cost-effectiveness or cost–benefit analysis based on intention to treat estimates (White, 2014).

Several other needs are also cited by researchers in the IE field. Many observers believe that there is a need for more unified and standardized outcome measures to make studies more comparable and informative (Kabeer et al., 2012; Stewart et al., 2012). In addition, although it has been suggested that program sustainability largely depends on the specific institutional environment, especially if the cash transfer or the service program is delivered at the local level, there is much less evidence on the most effective way to do this (Bouillon and Tejerina, 2007).

3. Evaluations and Policy Change

In this section, we first examine progress that the field has made in terms of taking IE results and findings and turning them into policy change. In the first subsection we present the view of the Evaluators (researchers conducting IEs in the field). In the second subsection we examine the view of the Implementers (aid agencies, international organizations and policymakers).

3.1. The View of the Evaluators

In the first part of this section we are interested in describing the view of the research community about the question: What is the state of impact evaluation—especially as it is affecting policy—a decade following the release of the Center for Global Development’s *When Will We Ever Learn?*

We started this exploration from the understanding that there is no doubt that the number of impact evaluations has skyrocketed over the past 10 years. In our interviews with some of the top IE researchers in the world, the primary questions that we were interested in understanding were twofold:

- Are the evaluations generating answers to questions of policymakers and donors?
- Are the findings generally in a format that these stakeholders can access and understand?

To carry out the survey, we sent surveys to 40 of the top IE scholars in the US and Europe and in other countries. We selected scholars from the major impact evaluation centers, including J-PAL, IPS and CEGD. We selected scholars from the major development banks and organization, including the World Bank, IDB and IFPRI. We also included individual scholars from universities, including economists and social scientists from Stanford, UCLA, San Diego, Princeton, Cornell, Oxford, Cambridge, LSE and more. Nearly half of the respondents answered our survey.

The survey consisted of eight questions. All of them gave the respondents a chance to assess changes over the past 10 years. The survey questions are included in the text below.

The survey also encouraged respondents to provide comments on top of the pre-coded survey choices. We also followed up with interviews with a subset of our respondents.

Question 1

[Does impact evaluation generate “bodies of evidence” that address the questions that policymakers, donors and implementers have?]

When asked about whether or not impact evaluation was generating “bodies of evidence” that address the questions that are of interest to policymakers, donors and implementers, evaluators had mixed ideas. While 30 percent of the respondents said unequivocally yes, half of the evaluators said “sometimes,” only 20 percent said outright “no.” In answering the question about changes over time, all respondents said that the IE literature was more focused on questions that were of interest to policymakers than 10 years before. Clearly, there seems to be a perception that researchers are putting more effort into policy-relevant work than they used to, although gaps still remain.

One of the respondents provided a comment that we believe explains in part the mixed results. The respondent said: “Look it. The literature is often, though not always, generating bodies of evidence that policymakers and donors need. Sometimes our research is producing more fundamental knowledge ... sort of like ‘basic research’ that may be useful in the longer run.” This seems to suggest that there remain perceived benefits to conducting research that is not of immediate value to policymakers and donors. Another interviewee said that to the extent that there was still a gap between what researchers do and what policymakers need, it is not always the fault of the research team. “In many cases they [policymakers] may not always realize that they need the evidence.” A number of our IE practitioners said that policymakers frequently do not question things enough. They often have preconceived ideas about what will and will not be effective and they are not really looking for evidence on something. However, by far and away, the older IE researchers made it clear that they definitely believe that it is more the case now than 10 years ago that bodies of policy relevant information are being created.

Question 2

[Always/Sometimes/Never: Before I start a new impact evaluation project or paper, I consult with policymakers or implementers about the questions that they are interested in.]

According to our data, most impact evaluation research teams do consult with policymakers about the questions that they are interested in. Indeed, about 40 percent of the respondents said they always did so; 50 percent said sometimes. Only ten percent of researchers said that they did not do so at all. However, in follow up interviews with the subset researchers that did not use consultations, several claimed that they read the literature closely enough that they believed they knew what policymakers were thinking from the literature.

In follow-up interviews it was found that although most researchers engage in consultations with policymakers before starting their research at least some of the time, the motivation for doing so differs. For example, one set of researchers suggested that the input of policymakers could be a valuable tool to increase the relevance of their research, but also argued in favor of maintaining some independence in research agenda. As one interviewee put it, “I definitely check what policymakers and implementers think/do/know but I sometimes decide that questions that they are not interested in are important nevertheless.”

Other researchers, whether interested in policy engagement or not, reporting finding that consultations are a necessary part of the research process. The idea is that since large scale IEs are often done through government channels, there has to be policymaker engagement just in order to get permission to do the

study in the first place. We were told by one researcher, “To do most impact evaluations requires partnering with government officials or implementers. As a result, understanding their needs and ensuring that the evaluation satisfies both mine and theirs is an important part of managing the relationship with the implementing partner.” If this is true more generally across settings and researchers, it would mean that the rise of IE is going to naturally be accompanied by more ex-ante consultations.

Finally, other researchers told us that many of these researcher-policymaker engagements are actually being driven by policymakers themselves. This is especially true when the government is the IE contractor. According to one researcher, “Often I am asked by a donor/policymaker to evaluate some program, so clearly I am doing something on a question that they are interested in. But I still sometimes do things that I am interested in and assume that donors/policymakers should be as well, but I do not consult with them.”

Hence, according to our findings, it appears that there is a lot of engagement between researchers and evaluators. However, it is not always being driven by the motivation to do research for the expressed purpose of policy change. In many cases, working on interesting evaluations requires consultations for getting permission to do the study in the first place and in other cases the rise of government interest in IE is creating opportunities for more of this sort of consultation.

Questions 3 and 4

[Always/Sometimes/Never: After I finish an impact evaluation project or paper, I disseminate those findings directly to policymakers or implementers.

and

If yes, how would you describe the format in which you disseminate findings?]

It is clear from our data and interviews that there is a trend by which researchers—whether or not they want to—are increasingly disseminating results to policymakers. Almost exactly half of the respondents said that they always did. The other half said sometimes. Even those that said they did not like doing so or believed that dissemination was difficult and time-consuming and not rewarded in the world of academia said that they were (sometimes) required to disseminate findings.

The dissemination from evaluators to policymakers is also found to be happening through multiple channels. Fully 80 percent of evaluators said they used workshops to disseminate results. Eighty percent also said that they used direct communications. Sixty percent said they produced formal or quasi formal briefs (or policy briefs) as one of the mediums for communications. Less, only 40 percent, used media outlets as a vehicle for dissemination.

From interviews, it seems that the experience of evaluators in undertaking their dissemination varies sharply. Some seem to imply it is a regular part of their post-analysis protocol. For example, one evaluator said, “I do it through IPA/JPAL as well as through in-country stakeholder workshops when relevant.” Another said: “I am in constant contact with the government unit that I am evaluating. I just send them regular emails that update them on the findings.”

Others find the process onerous. For example, one researcher told us:

“Sometimes. This is very hard. As an academic, I get little direct personal benefit from producing work that is accessible to policymakers or implementers. My incentives are set solely to publish academic papers in top journals. So, any effort that I put into this is either due to the fact that I personally value this kind of work or because it is necessary to secure the partnership of the implementing partner.”

Interviews found, however, that often times the dissemination is rather mechanical. As one researcher said, “I do it to ‘tick the box.’” Another said that he believed his work was being disseminated, but was not sure out effective the process was:

“I think the biggest source of this kind of work has been the development of organizations like the Jameel Poverty Action Lab and Innovations for Poverty Action which explicitly work to make the academic research of members more accessible. These groups usually write policy briefs or feature research on their websites largely using the time of their staff members. They provide very little input from researchers.”

Others do it because they are required to do so. According to our interviews, “Of course I do this for the evaluations that are done at the request of donors/policymakers. I do this more often now than 10 years ago because a greater share of my evaluations are done at the request of donors/policymakers.”

However, not all dissemination is being forced on researchers. We were told by more than one researcher that while dissemination to policymakers is not rewarded in the field, they continue to make the effort to do so out of personal commitment. One researcher said it plainly, “I got into academics to change the world. I go out of my way to try to make my research policy relevant and spend the time in communicating my results to policymakers. Not everyone is driven by this motivation. I am.”

In short, as in the case of consultations (see Question 2 above), there obviously is a lot more dissemination of research results and moving information from evaluator to policymaker than there was a decade ago. The flow of information is moving through many channels, formal and informal (and some through the media). Some (perhaps most) is being driven by requirements placed on researchers as direct or indirect IE contractual requirements. Others, however, are being pushed by the personal commitment of a subset of researchers.

Question 5

[Always/Sometimes/Never: Policymakers engage in dialogue about or seek to use results from my impact evaluation projects or papers.]

There were slightly less positive comments about the willingness of policymakers to become engaged in dialogue about IE results. While 30 percent said always and 40 percent said sometimes, 30 percent of the researchers responded never. Many also said there was a trend toward more dialogue over the past decade, but others said the trend was either slow or there was no trend at all.

Part of the problem seems to be policy salience. In other words, if the issue is hot, there is dialogue. If the issue is not hot, then there is less interest. According to one interviewee, “This [policy dialogue] only occurs if the paper is released amidst an existing policy debate. To be frank, other papers fall on deaf ears.”

There is also heterogeneity in responsiveness from policymakers and implementers. There were several interviewees that directly said policymakers were too busy or too uninterested to engage in serious dialogue. Another directly described the differences in those that were interested and those that were not, “Personally, I have received very little interest from policymakers about my work. The majority of the interest that I have received has been from foundations that support education programs or from groups like the World Bank.” Others are even more negative, “The political economy of policy dialogue and the role of impact evaluation is complex. Few policymakers have any interest.”

Questions 6 and 7:

[What are the benefits of doing work that is explicitly policy relevant?

and

What are the major barriers to doing policy relevant impact evaluations?]

There was a great deal of heterogeneity that appeared from our respondents in answering these questions. Thirty percent said doing policy relevant work helped with getting funding; 30 percent said it made getting funding more difficult. Forty percent of the respondents said that it helped get access to field resources (and access to study sites); the rest did not believe this was an issue. Only 10 percent said it made research easier to publish; more than half (60 percent) explicitly said it made research more difficult to publish. In fact, one said explicitly, “Does being policy relevant help during the publishing process? Very funny. This is absolutely not true.”

There were many other comments on the benefits of publishing and policy relevance. Some said that since they were committed to policy change, publishing papers that were policy relevant made his research agenda more interesting and important. Another said, “All research is policy relevant.” Several evaluators implied that it took effort to, “... strike a balance between doing something interesting academically and doing something that was policy relevant.”

Many of the same sentiments resounded when we asked evaluators about major barriers to doing work that is explicitly policy relevant. Many (40 percent) said that it was “time consuming” to work with policymakers. Others (also 40 percent) stated that researchers had to distort their research agenda to do policy-relevant work (often due to policymaker sensitivities). Only 20 percent said that it was absolutely unnecessary to distort the research agenda. As above, many (nearly half) said it took a toll on publishing. The toll is mostly, according to interviews, in the quality of publications. There is no reward in top journals for being policy relevant.

Question 8:

[Who, besides your own group, do you believe is best known for having policy impact?]

In answering this question, it was clear that the development banks, especially the World Bank (WB) and Interamerican Development Bank (IDB), were cited the most. Of those researchers that answered the question (about 30 percent did not), 75 percent cited either the WB or IDB as the most successful at using IE for policy impact. The International Growth Center was cited by 30 percent of respondents (conditional on giving a response).

The rest of the respondents cited both the large IE groups and individual researchers as having the most policy relevant program. The names JPAL, IPA and CEGA all were mentioned. Individual researchers were mentioned an equal number of times.

3.2. *The View of the Implementers*

In the second part of this section we are interested in describing the view of the policymaking and policy implementing community (henceforth, implementers) about the question: What is the state of the use of impact evaluation—especially as it is affecting the ways that they have used IE to guide policy, program and investments decisions?

Again, we started from the belief that there was no doubt that the number of impact evaluations had skyrocketed over the past 10 years. In our interviews with a wide range and large number of implementers, the primary questions that we are interested in understanding were twofold:

- Are the evaluations generating answers to questions that you, policy implementers, are interested in?
- Are the findings generally in a format that you, policy implementers, can access and find useful for policy, program and investment decisions?

To carry out the survey, we sent surveys to more than 50 policy implementers in both developing and developed countries. We surveyed three types of policy implementers, those in the development banking community (e.g., World Bank, etc), those in the bilateral funding community (e.g., USAID, DFID, etc) and actually policymakers (in Africa, India and Asia). The survey consisted of 8 questions. All of the questions gave the respondents a chance to assess changes over the past 10 years. The survey questions are included in the text below.

The survey also encouraged respondents to provide comments on top of the pre-coded survey choices.

Question 1:

[Think of a recent policy change you have been involved with. What was the primary mechanism behind that policy change? Did evidence play any role in that policy change?]

Implementers had a wide variety of answers to the question; evidence from impact evaluation was not the primary factor. More than half (54 percent) said the primary mechanism behind the recent policy change was the implementer's own initiative. They were acting on their own knowledge, experience and carrying out their assigned role. Political pressure was the mechanism identified in 23 percent of responses. Outside funders drove policy 8 percent of the time. Implementers cited evidence from impact evaluation as the primary mechanism only 15 percent of the time.

While evidence was not the *primary* mechanism behind policy change, most implementers did say that evidence did play some role in the roll out of the new policy. More than 60 percent of the respondents stated that evidence was a factor in the policy change. Around forty percent, however, said that evidence played no role whatsoever.

Question 2

[To the extent that evidence played a role where did come from?]

Interestingly when talking to policy implementers, there was a wide variety of answers to this question. Only about 15 percent of implementers said that the evidence came from researchers or research institutions per se. The most common answer (about 40 percent of the time) was that the evidence was generated as part of a project for which an IE had been specifically commissioned.

Also there was a lot of discussion in response to this question that evidence was not a one-time only proposition. Policy implementers often said that they needed to see information from many sources. If there was evidence from one national study, it was not enough. Evidence from multiple studies and from highly regarded international evaluation units (often, it should be noted, led by researchers from developed country universities) were more persuasive. According to respondents from development banks, the most persuasive evidence were those that came from within Bank meta-studies. If a single type of project was successful in creating impacts across multiple settings, more weight was given to it.

In summary, in the view of implementers, researchers—directly and indirectly (through in-project review teams)—have contributed to policy change. It is clearly not the driving mechanism. But, it seems to be playing an increasingly important role in influencing change.

Question 3

[Does impact evaluation generate “bodies of evidence” that address questions that you as a policymaker, donor or implementer have?]

The answers to this question were the most consistent of all questions. More than 90 percent of respondents said “yes,” IE generates evidence that addresses questions of implementers. Only 10 percent said “no.” Importantly, more than 75 percent of respondents said that IE was more important today in providing information that policymakers found useful than it was 10 years ago.

Question 4

[Always/Sometimes/Never: I have been consulted about the questions that are of interest to me (or my colleagues) before a researcher started a new IE project.]

There have been many more IEs over the past ten years. Our data shows that policy implementers are increasingly being consulted with before evaluators begin their research. However, they are not being consulted all of the time. According to the findings, in 77 percent of cases implementers say that evaluators sometimes seek consultations with them before beginning an IE project. Nearly two thirds say this is up significantly when compared to 10 years ago. Only 15 percent say this never happens (and 8 percent say always).

So, in summary, IE evaluators are making a lot more effort to consult with implementers. It is not 100% (by far). But implementers are clear in saying that it is much more common than it was a decade ago.

Question 5

[Always/Sometimes/Never: I have received findings directly from an impact evaluation researcher.

and

In what format did you receive the IE findings from researchers?]

The response to this question mirrors the previous one. Three quarters (75 percent) of implementers stated that they had “sometimes” received findings directly from an evaluator. They were clear that it was not always (only eight percent said “always”). However, only 17 percent said never.

Importantly, nearly 70 percent of the respondents said the availability of IE findings was up compared to a decade ago. No one said that there was less information than 10 years ago. Only 30 percent said it was the same.

According to our data, policymakers stated that they received IE findings in a variety of formats. Nearly 40 percent said they had attended workshops and heard the findings. More than half (54 percent) said they read policy briefs. Thirty eight percent said direct communications were the source of their access to IE findings. Only 30 percent said they used the media. Interestingly, the findings of this question are not that much different (in percentage terms) than the shares reported by evaluators to the same query (see above).

Question 6

[Always/Sometimes/Never: Researchers have engaged me in dialogue about the results from their IE projects.]

Nearly three quarters of implementers reported that they are either always (17 percent) or sometimes (58 percent) engaged with IE evaluators regarding the results of their IEs. Moreover, this is much more true now than earlier. Exactly two-thirds (67 percent) of respondents said that they are more likely to have a dialogue with an evaluator now than 10 years ago. Only one-third say it is the same as 10 years ago. No respondent said that dialogues were diminishing.

Question 7

[Always/Sometimes/Never: I (or my colleagues) have acted on the results from impact evaluation projects.]

Consistent with the results of other questions to implementers, nearly all (85 percent) of implementers stated that they had acted on the findings of IE projects. In interviews, respondents told us that evidence was important, but not the only thing that mattered when deciding to roll out a project.

When looking at the change over time, our data suggests that in nearly two thirds of cases, policymakers are acting on IE results. Only 30 percent of policy implementers said that they used results of IE projects no more than in the past.

Question 8

[Which IE group or individual researcher do you think is best known for having policy impact?]

Interesting, the groups identified as the best known IE evaluators are different for implementers than for evaluators. No policymaker mentioned the World Bank or IDB. The number one group was 3ie. Only a small minority stated J-PAL or IPA or CEGA. Interestingly, when asking the policymakers from developing countries, almost all of them cited local researchers.

4. Summary and Conclusions|

In this paper, we have shown that the rise in the quantity of IEs conducted over the past decade (documented earlier in the manuscript) has been at least partially matched by an increase in the generation of bodies of evidence that are relevant and accessible to policymakers and implementers. Nearly all evaluators and implementers agreed that impact evaluation findings were reaching policymakers (as measured by consultations, dissemination, and dialogue between evaluators and implementers) to a greater extent than they were before the release of *When Will We Ever Learn*.

While evaluators cited some remaining barriers to making evidence available to policymakers (such as incentives not to conduct directly policy-relevant work, little interest from policymakers, etc.), 90 percent of implementers reported that IE generates evidence that addresses the questions of implementers and 60 percent of implementers reported that evidence played a direct result in policy change. This upsurge in dialogue and exchange between evaluators and implementers is certainly something to be commended. However, our interviews also suggest that more work is needed: for instance, evaluators reported that incentives still cut against doing policy-relevant work and oftentimes dissemination is more by rote than carefully designed to be effective. Further work is therefore needed to make sure that engagement between evaluators and implementers can continue to increase both in quantity and in quality.

References

- Akresh, R., De Walque, D., and Kazianga, H. (2012). *Alternative cash transfer delivery mechanisms: impacts on routine preventative health clinic visits in Burkina Faso*. National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w17785>
- Barr, A., Dekker, M., and Fafchamps, M. (2012). Who shares risk with whom under different enforcement mechanisms? *Economic Development and Cultural Change*, 60(4), 677–706.
- Bouillon, C. P., and Tejerina, L. R. (2007). Do we know what works? A systematic review of impact evaluations of social programs in Latin America and the Caribbean. *A Systematic Review of Impact Evaluations of Social Programs in Latin America and the Caribbean (June 2007)*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=996502
- Clifford M, Miller T, Stasz C, Goldman C, Sam C, Kumar K (2013) How effective are different approaches to higher education provision in increasing access, quality and completion for students in developing countries? Does this differ by gender of students? A systematic review. RAND Corporation. ISBN: 978-1-907345-58-6.
- Dhaliwal, Iqbal, E. Duflo, R. Glennerster, and C. Tulloch. 2011. "Comparative Cost Effectiveness to Inform Policy in Developing Countries." Unpublished manuscript, Abdul Latif Jameel Poverty Action Lab, MIT.
- Evans, David and Popova, Anna, 2014. Cost Effectiveness Measurement in Development: Accounting for Local Costs and Noisy Impacts. World Bank Policy Research Working Paper 7027.
- Frymus, D., Kok, M., de Koning, K., Quain, E., 2013. Knowledge gaps and a need based Global Research agenda by 2015.
- Giedion, Ursula, Eduardo Andrés Alfonso, and Yadira Díaz. "The impact of universal coverage schemes in the developing world: a review of the existing evidence." Universal Health Coverage Study Series, No. 25, The World Bank (2013)
- Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel MJ Vermeersch. 2011. *Impact evaluation in practice*. World Bank Publications,
- Glewwe, P. Hanushek, E., Humpage, S., Ravina. R. "School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010." Eds. Paul Glewwe, Education Policy in Developing Countries, Chicago: University of Chicago Press, December 2013, pp. 13-64.
- Hafner, T., Shiffman, J., 2013. The emergence of global attention to health systems strengthening. *Health Policy Plan*. 28, 41–50. doi:10.1093/heapol/czs023
- Hagen-Zanker, J., McCord, A., Holmes, R., Booker, F., Molinari, E., and others. (2011). Systematic review of the impact of employment guarantee schemes and cash transfers on the poor. *London: Overseas Development Institute*. Retrieved from <http://re.indiaenvironmentportal.org.in/files/systematic-review.pdf>
- J-PAL. 2013. "J-PAL Youth Initiative Review Paper." Cambridge, MA: Abdul Latif Jameel Poverty Action Lab.
- Jamison, D.T., 2015. Disease Control Priorities, 3rd edition: improving health and reducing poverty. *The Lancet*. doi:10.1016/S0140-6736(15)60097-6
- Kabeer, N., Piza, C., and Taylor, L. (2012). What are the economic impacts of conditional cash transfer programmes: a systematic review of the evidence. *EPPI CENTRE, SOCIAL SCIENCE RESEARCH UNIT, INSTITUTE OF EDUCATION & UNIVERSITY OF LONDON (eds.)*. Retrieved from http://www.3ieimpact.org/media/filer_public/2014/10/17/naila_kabeer-_cts_systematic_review.pdf
- Kim, J. S., Samson, J. F., Fitzgerald, R., and Hartry, A. (2010). A randomized experiment of a mixed-methods literacy intervention for struggling readers in grades 4–6: Effects on word reading

- efficiency, reading comprehension and vocabulary, and oral reading fluency. *Reading and Writing*, 23(9), 1109–1129.
- Kingdon, G., Aslam, M., Rawal, S., and Das, S. (2012). Are contract and para-teachers a cost effective intervention to address teacher shortages and improve learning outcomes?
- McEwan, P.J., 2012. Cost-effectiveness analysis of education and health interventions in developing countries. *Journal of Development Effectiveness* 4, 189–213.
doi:10.1080/19439342.2011.649044
- McEwan, P.J., 2014. Improving Learning in Primary Schools of Developing Countries A Meta-Analysis of Randomized Experiments. REVIEW OF EDUCATIONAL RESEARCH 0034654314553127.
doi:10.3102/0034654314553127
- Meyer, C., University of London, Social Science Research Unit, Evidence for Policy and Practice Information and Co-ordinating Centre, 2011. The impact of vouchers on the use and quality of health goods and services in developing countries a systematic review. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, London.
- Miller, C. M. (2011). Cash transfers and economic growth: a mixed methods analysis of transfer recipients and business owners in Malawi. *Poverty and Public Policy*, 3(3), 1–36.
- Morgan, C., Petrosino, A., and Fronius, T. (2013). A Systematic Review of the Evidence of the Impact of School Voucher Programmes in Developing Countries. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Morgan, C., Petrosino, A., and Fronius, T. (2014). Eliminating School Fees in Low-Income Countries: A Systematic Review. *Journal of MultiDisciplinary Evaluation*, 10(23), 26-43.
- Murnane, R. J., and Ganimian, A. J. (2014). Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations (No. w20284). National Bureau of Economic Research.
- Radermacher, R., Dror, I., Zabel, K., and Ashok, S. (2009). The impact of microinsurance on economic and social outcomes for the poor in low-and middle income countries. Retrieved from http://campbellcollaboration.org/lib/download/686/Radermacher_Microinsurance_Title.pdf
- Toikka, R., Gais, T., Nikolov, P. V., and Billen, P. (2004). *Spending on social welfare programs in rich and poor states*. EconWPA. Retrieved from <https://ideas.repec.org/p/wpa/wuwppe/0410006.html>
- Rockers, P.C., Bärnighausen, T., 2013. Interventions for hiring, retaining and training district health systems managers in low- and middle-income countries, in: *Cochrane Database of Systematic Reviews*. John Wiley and Sons, Ltd.
- Rowe, Alexander K., Samantha Rowe, David H. Peters, Kathleen A. Holloway, John Chalker, Dennis Ross-Degnan. 2014. Health Care Provider Performance Review: Systematic Review of Strategies to Improve Health Care Provider Performance in Low and Middle Income Countries. Paper Presented at the Workshop on the Quality of Health Services and Products in Developing Countries, World Bank.
- Stewart, R., Van Rooyen, C., Korth, M., Chereni, A., da Silva, N. R., and De Wet, T. (2012). *Do Micro-credit, Micro-savings and Micro-leasing Serve as Effective Financial Inclusion Interventions Enabling Poor People, and Especially Women, to Engage in Meaningful Economic Opportunities in Low-and Middle-income Countries?: A Systematic Review of the Evidence*. EPPI-Centre.
- Swain, R. B. (2012). Differential impact of microfinance delivery mechanism on vulnerability. *Applied Economics Letters*, 19(8), 721–724.
- Tripney, J. S., and Hombrados, J. G. (2013). Technical and vocational education and training (TVET) for young people in low-and middle-income countries: a systematic review and meta-analysis. *Empirical Research in Vocational Education and Training*, 5(1), 1-14.
- White, H. (2014). Current challenges in impact evaluation. *European Journal of Development Research*, 26(1), 18–30.

Willey B, Smith Paintain L, Mangham-Jefferies L, Car J, Armstrong Schellenberg J (2013)
Effectiveness of interventions to strengthen national health service delivery on coverage, access,
quality and equity in the use of health services in low and lower middle income countries.
London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of
London.

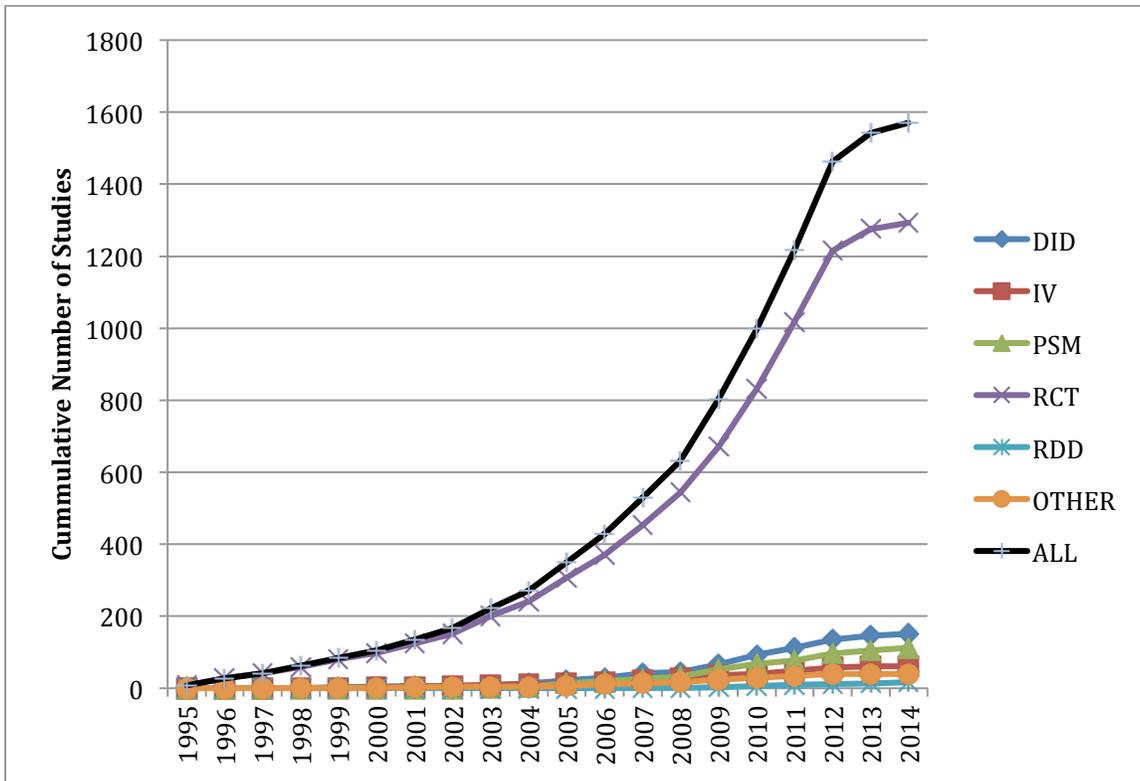


Figure 1: Number of Health-Oriented Impact Evaluation Studies in Developing Countries (Experiments and Quasi-Experiments) Over Time

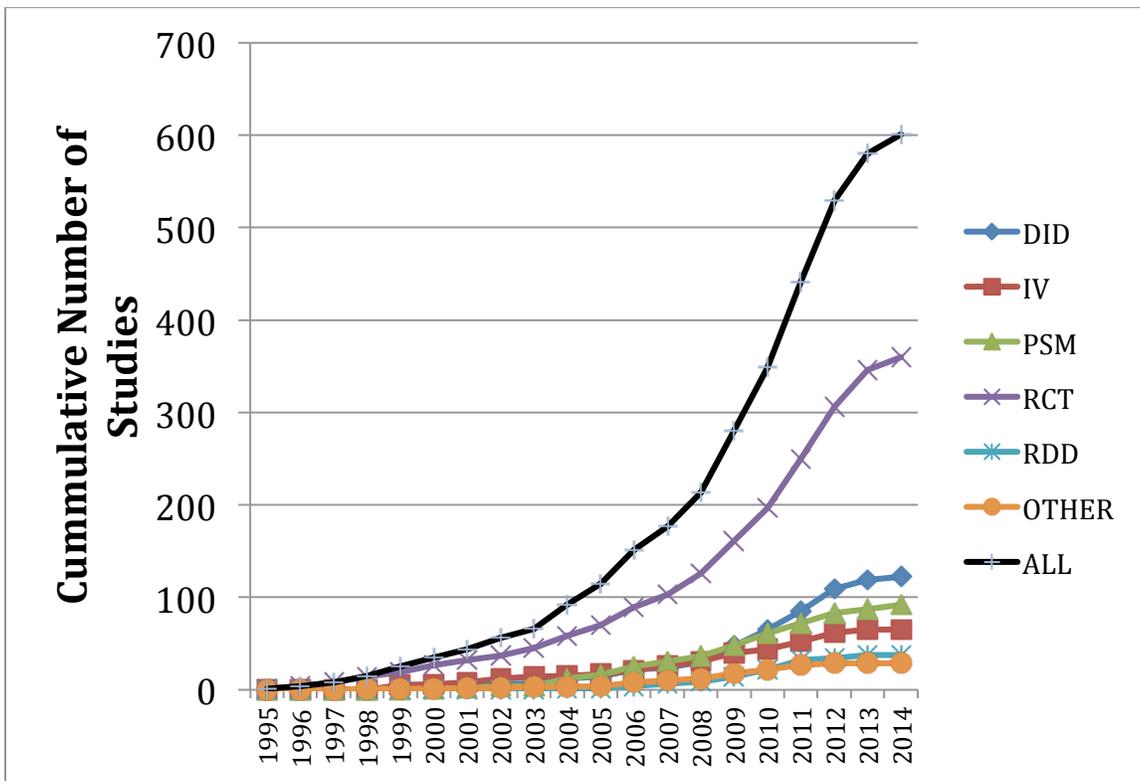


Figure 2: Number of Education-Oriented Impact Evaluation Studies in Developing Countries (Experiments and Quasi-Experiments) Over Time

Table 1: Growth in Health-Focused Impact Evaluation Studies in Developing Countries over Two Decades (by Health Area)

	Pre 2004	2005-2014	Difference	% Change
Child Nutrition	83	231	148	178%
Nutrition	301	1296	995	331%
Health Sector Reform	5	30	25	500%
Health Financing, Insurance and User Fees	6	45	39	650%
Health Services	191	804	613	321%
HIV/AIDS	39	233	194	497%
Hospitals - Secondary & Tertiary	10	42	32	320%
Mortality	17	88	71	418%
Preventative Health and Health Behavior	235	998	763	325%
Primary Health (incl. reproductive health)	58	223	165	284%
Sexual Behavior	33	169	136	412%
Specific Diseases (incl. Malaria and TB)	96	281	185	193%

Data source: 3ie impact evaluation database (<http://www.3ieimpact.org/en/evidence/impact-evaluations/>)

Table 2: RCTs as Percentage of Total Health Impact Evaluations in Developing Countries Between 1995 and 2014

	RCTs 1995- 2004	RCTs 1995- 2004 (%)	RCTs 2005- 2014	RCTs 2005- 2014 (%)	Percentage Point difference
Child Nutrition	71	86%	186	81%	-5%
Nutrition	269	89%	1049	81%	-8%
Health Sector Reform	3	60%	16	53%	-7%
Health Financing, Insurance and User Fees	1	17%	12	27%	10%
Health Services	186	97%	716	89%	-8%
HIV/AIDS	37	95%	211	91%	-4%
Hospitals - Secondary & Tertiary	9	90%	37	88%	-2%
Mortality	15	88%	64	73%	-16%
Preventative Health and Health Behavior	220	94%	866	87%	-7%
Primary Health (incl. reproductive health)	53	91%	177	79%	-12%
Sexual Behavior	30	91%	145	86%	-5%
Specific Diseases (incl. Malaria and TB)	94	98%	259	92%	-6%

Data source: 3ie impact evaluation database (<http://www.3ieimpact.org/en/evidence/impact-evaluations/>)

Table 3: Growth in Education-Focused Impact Evaluation Studies in Developing Countries over Two Decades (by Education Area)

Education Area	1995-2004	2005-2014	Difference	% Change
Distance education/Education technology	2	12	10	500%
Educational inputs	12	73	61	508%
Girls' education	4	27	23	575%
Non-formal Education	3	25	22	733%
Pre-primary and Primary Education	45	203	158	351%
Public/Private Sector Education	6	20	14	233%
Secondary Education	20	99	79	395%
Student Loan	1	3	2	200%
System Reform & Capacity Building	9	62	53	589%
Tertiary Education	1	25	24	2400%
VET	6	74	68	1133%
Scholarships	4	20	16	400%

Data source: 3ie impact evaluation database (<http://www.3ieimpact.org/en/evidence/impact-evaluations/>)

Table 4: Growth in Social Policy Impact Evaluation Studies in Developing Countries over Two Decades (by Social Policy Area)

Theme area	All studies 1995-2004	All studies After 2004	RCTs 1995-2004	RCTs After 2004
	[1]	[2]	[3]	[4]
[1] Conditional/Unconditional cash transfer	28	190	19	104
[2] Microfinance	10	148	1	65
[3] Employment and labor market	32	202	2	71
[4] Social security/safety net	3	18	0	3
[5] Insurance	9	87	2	33
[6] Pension	2	4	0	0
[7] Total	84	649	24	276

Data source: 3ie impact evaluation database (<http://www.3ieimpact.org/en/evidence/impact-evaluations/>)