



UNITED NATIONS
UNIVERSITY
UNU-WIDER

WIDER Working Paper 2020/20

The impact of impact evaluation

Are impact evaluation and impact evaluation synthesis contributing to evidence generation and use in low- and middle-income countries?

Richard Manning,¹ Ian Goldman,² and Gonzalo Hernández Licona³

March 2020

Abstract: In 2006 the Center for Global Development’s report ‘When Will We Ever Learn? Improving lives through impact evaluation’ bemoaned the lack of rigorous impact evaluations. The authors of the present paper researched international organizations and countries including Mexico, Colombia, South Africa, Uganda, and Philippines to understand how impact evaluations and systematic reviews are being implemented and used, drawing out the emerging lessons. The number of impact evaluations has risen (to over 500 per year), as have those of systematic reviews and other synthesis products, such as evidence maps. However, impact evaluations are too often donor-driven, and not embedded in partner governments. The willingness of politicians and top policymakers to take evidence seriously is variable, even in a single country, and the use of evidence is not tracked well enough. We need to see impact evaluations within a broader spectrum of tools available to support policymakers, ranging from evidence maps, rapid evaluations, and rapid synthesis work, to formative/process evaluations and classic impact evaluations and systematic reviews.

Key words: evaluation, evidence use, impact evaluation, research synthesis, systematic reviews

JEL classification: D04, O12, P45

Acknowledgements: The authors acknowledge the contributions from Tony Addison at UNU-WIDER, who commissioned the study and provided detailed inputs. We also acknowledge the authors of the short country contributions: Lucas Gómez and Clara Lorena Trujillo from Colombia; Carlos Bernardo O. Abad Santos, Violeta S. Corpus, and William C. Ku from Philippines; and Abdul Muwanika from Uganda. We would further like to acknowledge Judy Scott-Goldman, who assisted with editing, as well as Thania de la Garza and Janet Zamudio, who helped with their views and knowledge of the Mexican case. We are also grateful for the time and attention given by our various interlocutors listed in Appendix B.

¹ Blavatnik School of Government, University of Oxford, UK, corresponding author: richardgmanning@btinternet.com; ² CLEAR Anglophone Africa, Parktown, South Africa; ³ Instituto Tecnológico Autónomo de México (ITAM), Mexico City, Mexico

This study has been prepared within the UNU-WIDER project on [Extractives for development \(E4D\)](#), which is part of a larger research project on [Macro-economic management \(M-EM\)](#).

Copyright © UNU-WIDER 2020

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-777-4

<https://doi.org/10.35188/UNU-WIDER/2020/777-4>

Typescript prepared by Joseph Laredo.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

1 Introduction

1.1 Why is this issue worth studying?

In 2006, the Center for Global Development published a report entitled ‘When Will We Ever Learn? Improving lives through impact evaluation’, which criticized the international donor community in particular for failing to build a stronger evidence base on effective social development programmes, notably by failing to ensure regular, rigorous evaluations of the impact of such programmes (see Box 1).

The report argued for a new collective push to change the situation, and concluded:

Will we really know more in 10 years? [...] The international community could be in one of two situations.

We could be as we are today, bemoaning the lack of knowledge about what really works and groping for new ideas and approaches to tackle the critical challenges of strengthening health systems, improving learning outcomes, and combating the scourge of extreme poverty.

Or we could be far better able to productively use the resources for development, based on an expanded base of evidence about the effectiveness of social development strategies.

Which of those situations comes to pass has much to do with the decisions that leaders in developing country governments, NGOs, and development agencies make over the next couple of years about conducting impact evaluations (2006: 42–43).

Box 1: Definitions

The Organisation for Economic Co-operation and Development’s Development Assistance Committee (OECD-DAC) defines evaluation as ‘the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should also provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors’.

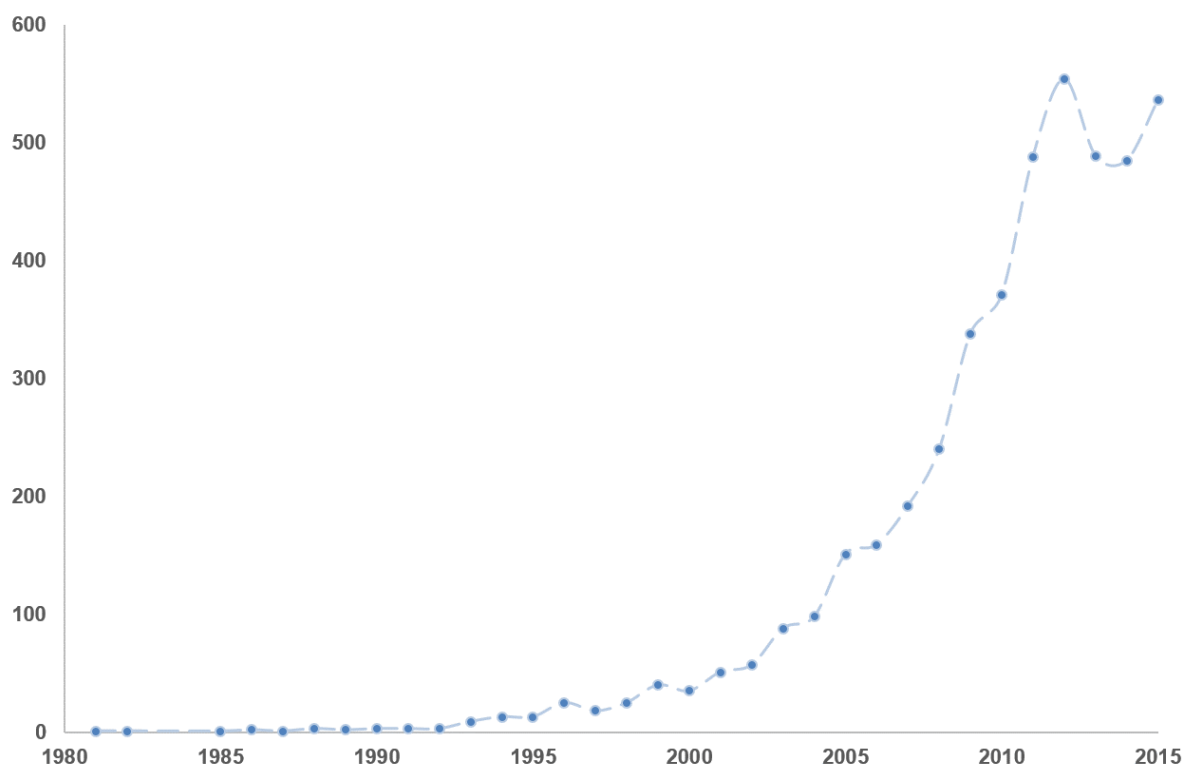
The OECD defines impact evaluation as ‘an assessment of how the intervention being evaluated affects outcomes, whether these effects are intended or unintended’ (OECD 2006: 1). It also quotes the IFAD impact evaluation guidelines, which define impact as the ‘the attainment of development goals of the project or program, or rather the contributions to their attainment’ (ibid.). In this case impacts are defined as ‘positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended’ (OECD 2002: 24). In this review we also consider products that synthesize impact evaluations, notably systematic reviews and evidence maps, sometimes referred to as evidence gap maps (EGMs).

A key element in impact evaluation is the concept of causality: that the outputs and outcomes result in the desired or unintended impacts.

On the face of it, huge progress has been made since 2006. Annual publication of impact evaluations (IEs) increased dramatically after 2008, as set out in a study by Cameron, Mishra, and Brown (2015) of the 2,259 IEs included in the IE Repository (IER) of the International Initiative for Impact Evaluation (3ie) that were published between 1981 and 2012.

A subsequent analysis of the data on IEs in the IER, to September 2015, when there were 4,205 studies in the IER,¹ shows the following picture of completed studies by year:

Figure 1: Numbers of impact evaluations completed, by year of completion



Source: IER (courtesy of 3ie).

In addition, there has been a parallel growth in systematic reviews (SRs) and other syntheses of the findings of this growing body of work,² and more recently in evidence maps, which assess where rigorous evidence exists or is absent.

This growth in the production of IEs has been seen in individual countries; in specific programmes focused on impact evaluation, such as the World Bank’s Development Impact Evaluation (DIME) and Strategic Impact Evaluation Fund (SIEF); in initiatives such as the Abdul Latif Jameel Poverty Action Lab (J-PAL), Innovations for Poverty Action (IPA), the Center for Effective Global Action (CEGA), and 3ie (whose creation was a direct response to the CGD report); in self-funded initiatives within development agencies, often quite decentralized, or in governments; and in the academic sector, where quantitative analyses of impact have become a staple of economic and social research (Banerjee et al. 2017). While this paper was in draft form, the 2019 Nobel Prize in Economics was awarded to three leading exponents of impact evaluation—Abhijit Banerjee,

¹ Sabet and Brown (2018).

² ‘In international development there were few reviews before 2008, after which the number grew steadily to over 100 published in 2016.’ (White 2019, quoting the 3ie database).

Esther Duflo, and Michael Kremer— ‘for their experimental approach to alleviating global poverty’.³

This paper attempts to ascertain what the impact of this growth has been. We examine, in particular, the extent of adoption of these techniques; the impact on wider ‘evidence systems’ in the countries concerned and on evaluation generally; and the use made of findings by policymakers in the countries of implementation, in relevant international policy communities, and in development agencies. We also consider what factors appear to encourage or weaken such use. One conclusion from our work is that among both countries and development agencies there is a wide disparity of interest and practice in the generation and use of rigorous evidence of impact. This provides, in our view, a valuable opportunity to reflect on what approaches provide lessons from which others can usefully learn.

As with the CGD report, we concentrate on impact evaluations of interest to low- and middle-income countries (LMICs) and to the international development community,⁴ rather than the wider use of such evaluations by ‘rich’ countries. We accept that many factors are likely to be common to countries at every level of development. But the fact that a significant proportion (in some countries, an overwhelming proportion⁵) of impact evaluations are financed and/or commissioned by development agencies rather than by governments or other local actors makes the situation distinctively different, particularly in the more aid-dependent countries. Bilateral development agencies have historically been ahead of many other agencies of their own governments in promoting the use of evaluations of all types, and some might argue that they can on occasion push poorer countries to achieve standards of monitoring and evaluation that their own administrations do not routinely meet.

We have also seen arguments that the preponderance of donor funding and commissioning might raise risks of skewing evaluations (not just impact evaluations), along with other forms of research, towards positive findings of donor-supported programmes—see, notably, the critique of a group of health experts that researchers’ ‘involvement in a collective drive to demonstrate success’ can unintentionally ‘instil a fear of failure, stifle risk-taking and innovation, and lead to the fabrication of achievement’. They also assert that ‘perverse incentives exist across the global health and development sectors to use simplistic indicators of success and bad or fudged data’ (Storeng et al. 2019).

Of course, safeguards can be and are put in place to preserve objectivity.⁶ We also accept that there is a case for a degree of ‘supply push’ by donors in order to create a critical mass of impact

³ The citation asserts that ‘Their experimental research methods now entirely dominate development economics’. Press release: The Prize in Economic Sciences 2019. Available at: <https://www.nobelprize.org/prizes/economic-sciences/2019/press-release/> (published 14 October 2019)-

⁴ Unlike the CGD report, we do not limit ourselves to the social sectors. These remain predominant, but the use of similar techniques for assessing impact has grown quite encouragingly in other sectors.

⁵ For example, in Uganda (see Section 3.4 below).

⁶ For example, J-PAL have guidance on pre-registration, data transparency and pre-analysis plans. Available at: <https://www.povertyactionlab.org/research-resources/transparency-and-reproducibility> (accessed 7 November 2019).

evaluations, without which it is difficult to know how IEs can help, and impossible to produce useful SRs or other synthesis products.⁷

Nevertheless, we believe that evaluation results would be better grounded and that evaluation systems could be more robust if more evaluations were commissioned/financed by the countries themselves. Experiences in India, China, Mexico, South Africa, Uganda, Philippines, and Colombia suggest that if more impact evaluations are to be commissioned by countries themselves, three things are of particular importance: the demand for transparency, for example from civil society and opposition parties; the recognition by efficient governments of the need to find concrete technical policy solutions; and adequate evaluation capacities by local researchers.

Our interest in the topic of this paper arises from our experience. Two of us have led important initiatives in our countries to generate and use rigorous evidence of the success or otherwise of national programmes, and one has worked in international development and was a participant in the CGD study. All three of us have served on the Board of 3ie. It must, however, be stated that this paper is not a 3ie study.

We share a view that rigorous evidence of the impact of programmes is a valuable part of an evidence system, which needs to be built on reliable and accessible data, regular monitoring, and a high-level commitment to promote the use of evidence to improve policy and practice. Not every initiative lends itself to the common IE toolbox of randomized control trials and experimental designs, but these approaches are extremely important for assessing the impact of particular initiatives in a rigorous way. However, our experience is that use of such evidence seldom follows a simple linear path, and will reflect many contextual factors.

We are, of course, conscious of a counter-narrative: the denigration in some quarters of evidence, particularly where it may raise questions about programmes with strong government commitment, and a broader distrust of ‘experts’. We hope that our paper will offer some useful reflections on both the adoption and the use of more rigorous techniques for assessing impact, and on how to strengthen the prospects that lessons will indeed be learned from the growing body of evidence on international development, including through a clearer vision of IE as one important part of a wider evidence system.

1.2 Methodology

We have not been able to find any independent study that looks in the round at our core question, ‘Are impact evaluation and impact evaluation synthesis contributing to evidence generation and use in low- and middle-income countries?’ In particular, studies that use experimental techniques to answer this question even for a single policy are scarce,⁸ and there are few independent evaluations of either development agency- or country-based IEs.⁹

⁷ Chris Whitty (personal communication), who notes that this was also true in developing useful and usable SRs in medicine, and argues that ‘Persisting with a supply of high-quality IEs and SRs before there is demand is essential if they are likely to have any chance of taking off.’

⁸ For an exception, see Hjort et al. (2019).

⁹ Three examples are Crespo and Azuara Herrera (2017) on the Inter-American Development Bank, Goldman et al. (forthcoming) on South Africa, and Ramirez et al. (2012) on the World Bank.

However, there is a good deal of evidence available from individual agencies specializing in impact evaluation, such as J-PAL, IPA, CEGA, and 3ie;¹⁰ from research and evaluation departments in many agencies both among assistance providers and increasingly from the countries themselves; and from those actually carrying out IEs and synthesis products. Much of this evidence is about the effects of individual IEs or synthesis products,¹¹ but there are also claims of impact for either groups of studies on specific areas of policy¹² or of the body of research on the research industry itself (Banerjee et al. 2017).

We have therefore assembled five case studies of LMICs with the assistance of colleagues from those countries, and contacted a wide variety of agencies that finance, commission, and/or execute IEs and, in some cases, synthesis products. Appendix B lists all those to whom we have spoken or who have responded to requests for information. It was clear from our discussions that there is a serious interest in continuing to build essential evidence of impact while also improving the use of such evidence.

For our countries, we selected Mexico, Colombia, South Africa, Uganda, and Philippines because we wanted to cover various regions and include public systems where evaluation processes have been established and IEs have had some time of exposure in order to see potential impacts on public policy, especially on policymakers. We acknowledge that addressing more country cases and examples of NGOs—not least Southern-based NGOs such as BRAC, which works with IEs—is a pending issue for another paper.

In order to structure feedback, we developed the checklist shown in Appendix C. This was filled in by some agencies and used as a basis for semi-structured interviews with others. We have quoted several comments that seem to us to have particular resonance, and we have drawn heavily on the overall body of feedback that we received. We also shared an advance draft of this paper with a number of our informants to check how far our emerging findings did or did not accord with their own experience, and have made a number of adjustments in the light of this feedback. However, our observations are our own, except where stated to the contrary.

1.3 How to frame the use of evaluations

In terms of analysing the use of research evidence, the classic text is from Weiss (1979). Weiss suggests that the results of basic research may be used for applied research, development, and application: ‘basic research discloses some opportunity that may have relevance for public policy; applied research is conducted to define and test the findings of basic research for practical action; if all goes well, appropriate technologies are developed to implement the findings; whereupon application occurs.’ She calls this the *Knowledge-driven model*. She also refers to a *Problem-solving model*, which involves the ‘direct application of the results of a specific social science study to a pending decision’ or the purposeful commissioning of social science research and analysis to fill knowledge gaps. A third type of use is the *Interactive model*, where ‘All kinds of people involved in an issue area

¹⁰ E.g. ‘Evidence uptake and use from 3ie-supported work: Narrative summaries of examples’. 3ie, March 2019. Unpublished (personal communication; hereafter 3ie (March 2019)).

¹¹ E.g. Kingra and Leach (2019) on cash transfers in Malawi.

¹² E.g. ‘Microcredit: Impacts and limitations’, https://www.povertyactionlab.org/full-search?search_api_views_fulltext=Microcredit:%20Impacts%20and%20Limitations%E2%80%99 (accessed 9 September 2019).

pool their talents, beliefs, and understandings in an effort to make sense of a problem'.¹³ A fourth type is the *Political model*, where policymakers have taken a 'stand that research is not likely to shake. In such cases, research can still be used. It becomes ammunition for the side that finds its conclusions congenial and supportive'. Finally, the *Enlightenment model* considers 'social science generalizations and orientations percolating through informed publics and coming to shape the way in which people think about social issues' (Weiss 1979).

The use of evaluation findings is commonly described as *instrumental, conceptual, process, or symbolic use*. Johnson et al. (2009) define these terms as follows:

Instrumental use refers to instances where someone has used evaluation knowledge directly. Conceptual use refers to cases when no direct action has been taken but where people's understanding has been affected. Symbolic use refers to examples where a person uses the mere existence of the evaluation, rather than any aspect of its results (2009: 378).

Patton defines process use as 'individual changes in thinking and behaviour and program or organizational changes in procedures and culture that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process' (1997: 90).

Table 1 provides examples of the types of evaluation use in the South African context.

Table 1: Examples of types of evaluation use in South Africa

National department	Type of use
Basic Education (DBE)	<i>Instrumental use</i> —As a result of the Funza Lushaka (teacher bursary) evaluation, DBE's budget for the management information system was increased. The Funza Lushaka bursary scheme now also has Cabinet approval to appoint new staff.
Human Settlements (DHS)	<i>Conceptual use</i> —Respondents from DHS noted that they find evaluations helpful as a reflective experience, but not currently as a decision-making exercise.
Justice and Cooperative Development (DJCD)	<i>Process use</i> —Respondents noted that being part of an evaluation in the National Evaluation Plan was helpful in that it shed light on good practices in the evaluation process and influenced their internal processes. This opens up the possibility that more evaluations can be managed internally. Like other departments and provinces, respondents in DJCD found the theory of change to be the most useful evaluative mechanism.
Department of Health	<i>Symbolic use</i> —A national evaluation of nutrition interventions for children under five in South Africa provides an example of symbolic use in that the evaluation raised the profile of the issue of child malnutrition (reference).

Source: adapted from Goldman et al. (forthcoming).

It is increasingly clear that the relationship between research and impact is indirect, non-linear, and more complex than evaluation frameworks typically allow (Reed et al. 2018). Reed also points out that benefits arising from research may be long-term and not simply instrumental. This is likely to be true of evaluations too, but as they relate to specific projects, programmes, and policies it may be easier to apply the results instrumentally.

So, what are the factors that facilitate or impede evidence use? Oliver (2014) undertook a systematic review that indicated the following barriers to, and facilitators of, research use.

¹³ Other authors point to the importance of relationships (e.g. Stewart et al. 2018) and social networks (e.g. Reed et al. 2018).

Table 2: Facilitators/barriers to research use

Top 5 barriers to use of evidence	Top 5 facilitators of evidence use
Availability and access to research/improved dissemination (n = 63 studies)	Availability and access to research/improved dissemination (n = 65)
Clarity/relevance/reliability of research findings (n = 54)	Collaboration (n = 49)
Timing/opportunity (n = 42)	Clarity/relevance/reliability of research findings (n = 46)
Policymaker research skills (n = 26)	Relationship with policymakers (n = 39)
Costs (n = 25)	Relationship with researchers/info staff (n = 37)

Source: Oliver et al. (2014: table 1), under a creative commons license (CC BY 2.0).

In 2016, Langer et al. undertook a systematic review of the barriers to, and facilitators of, research uptake. This indicated that there was *evidence of use* when (:

- Interventions facilitate[e] *access* to research evidence, for example through communication strategies and evidence repositories, conditional on the intervention design simultaneously trying to enhance decision-makers' opportunities and motivation to use evidence (reliable evidence).
- Interventions [build] *decision-makers' skills* to access and make sense of evidence (such as critical appraisal training programmes), conditional on the intervention design simultaneously trying to enhance both *capability* and *motivation* to use research evidence (reliable evidence).
- Interventions [foster] *changes to decision-making structures and processes* by formalising and embedding one or more of the other mechanisms of change within existing structures and processes (such as evidence-on-demand services integrating push, user-pull and exchange approaches) (cautious evidence) (Langer et al. 2016: 1).

By contrast, there appeared to be *no effect* from:

- Interventions that take a *passive approach to communicating evidence*[, which] only provide opportunities to use evidence (such as simple dissemination tools) (reliable evidence).
- Multi-component interventions that take a *passive approach to building evidence-informed decision-making (EIDM) skills* (such as seminars and 'communities of practice' without active educational components) (cautious evidence).
- Skill-building interventions applied at a *low intensity* (such as a one-off, half-a-day capacity-building programme) (cautious evidence).
- Overall, *unstructured interaction and collaboration* between decision-makers and researchers tended to have a lower likelihood of success. However, clearly defined, light-touch approaches to facilitating interaction between researchers and decision-makers, engagement in particular, were effective in increasing intermediate outcomes (cautious evidence) (Langer et al. 2016: 2).

The next section draws on reflections by major financiers and suppliers of impact evaluations and systematic reviews.

2 Overview of adoption and use of IEs and syntheses of IEs by major international financiers/suppliers)

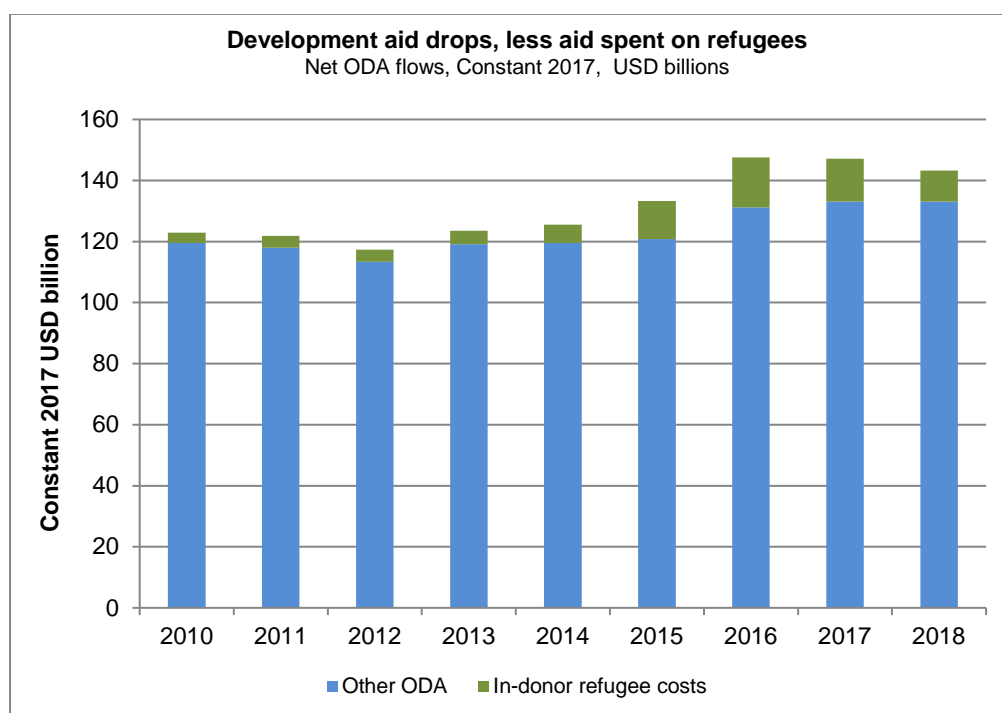
2.1 Overview of adoption of IEs and syntheses of evaluations

Some aggregate data on production and financing

As already mentioned, IEs in LMICs have grown significantly since the early years of the century,¹⁴ though, as shown in Figure 1, with a levelling off after 2011.¹⁵ It should also be noted that they remain a small share of evaluations, whether those funded by donor agencies or those funded by LMICs themselves.¹⁶

A significant—possibly in some cases dominant—proportion of IEs and synthesis products in respect of LMICs is still donor-financed. As Figure 2 shows, overall official development assistance (ODA), as recorded by the OECD Development Assistance Committee, has levelled off in real terms since 2010, when expenditure on refugees in donor countries is removed from the statistics.

Figure 2: Trend in official development assistance net flows, 2010–2018



Source: OECD (2019); DAC statistics. Adapted with permission.

¹⁴ For example, the production of IEs at the World Bank Group grew from an average of 16 initiated per year in the period 1999–2004 to an average of 57 per year in 2005–2010 (IEG 2012).

¹⁵ 3ie will be producing a new analysis of the IER in the first half of 2020.

¹⁶ For donors, Raifman et al. (2018) calculate that, from a sample of 299 evaluations commissioned by five donor agencies active in the field of health, only 7 per cent were IEs. In other sectors, the proportion could well be lower. CONEVAL, one of the largest commissioners of evaluations among LMICs, has commissioned over 2,800 evaluations since 2007, but only some 11 IEs (see Section 3.1).

Although expenditure on IEs is trivial in relation to total ODA, it may well be that the period of rapid expansion of official donor funding for impact evaluations has indeed ended. The data provided by funding agencies do not, unfortunately, enable us to be certain about this finding.

On the other hand, other sources are likely to be expanding. The production of randomized controlled trials (which continue to account for the majority of IEs) has shown robust growth and widening sector coverage, much of it financed by research grants and own resources, as reported by Banerjee et al. (2017). Foundations may still be expanding their financing, if, in many cases, from a low base. Some middle-income countries and a smaller number of low-income countries are certainly doing so from their own resources, usually also from a very low base (see Section 2.1.2). As the experience of the Inter-American Bank shows,¹⁷ borrowers may also be ready to include IEs in project finance from established multilateral sources.

However, a closer look at the sources of finance for IEs and associated synthesis products shows a very significant concentration of funding and commissioning among both official agencies and foundations. Most bilateral agencies in member countries of the OECD-DAC, for example, commission either a very small number of IEs a year, or none at all. At the other end of the scale, USAID has produced some 126 IEs since 2012 (until the end of the 2018 financial year, 30 September 2018).¹⁸ Among multilaterals, there is also great variety in commissioning, as well as in the extent to which agencies access bilateral funding (see below). Among foundations, the Bill and Melinda Gates Foundation and the Hewlett Foundation stand out as major financiers of IEs. The Bill and Melinda Gates Foundation is by far the largest contributor in absolute terms, though several other North American Foundations are involved,¹⁹ as are the UK-based Children's Investment Fund Foundation (CIFF) and Wellcome Trust.

It is a similar story with bilateral contributions to the larger multilateral programmes for IEs. Under the World Bank's DIME programme, 145 out of its total portfolio of 151 studies received seed funding from the i2i Trust Fund,²⁰ most of which is financed by a £20.5m contribution from DFID. Its SIEF programme, which has produced some 80 IEs, after initial funding of US\$10.4m from Spain in 2007, is currently funded by DFID (£30.6m) and CIFF (US\$3m). 3ie, which has produced 128 IEs as well as some formative evaluations, has for most of its history depended heavily on finance from DFID, the Bill and Melinda Gates Foundation, and the Hewlett Foundation. 3ie has worked hard to widen its funding base, but has found it difficult to secure comparable funding from other official sources or foundations, and has had to accept a significant switch from core funding to funding for specific initiatives. Donors do, however, sometimes use multilateral replenishment exercises to incentivize the production of IEs through the resources of the institution in question.²¹

¹⁷ Crespo and Azuara Herrera (2017).

¹⁸ Tania Alfonso (personal communication). Information on the number of IEs and related products is often poorly documented at bilateral agency level (the Millennium Challenge Corporation, along with USAID, being an exception, with 37 IEs since start-up). This is often because the commissioning of evaluations, including IEs, is largely decentralized. We were surprised to discover, for example, that DFID was unable to provide specific figures, though directing us to many interesting IEs and synthesis products.

¹⁹ For example, the funders of J-PAL include the Laura And John Arnold Foundation, the John D. and Catherine T. MacArthur Foundation, and the Alfred P. Sloan Foundation, as well as the Saudi Arabian-based Community Jameel.

²⁰ Science for Impact, World Bank Group (Legovini et al. 2019). Other contributors to the i2i Trust Fund are the European Union, the Norwegian Agency for Development Co-operation, and the World Food Programme.

²¹ For example, IFAD agreed to 17 IEs of projects under its Tenth Replenishment (2015–2017).

Financing of IEs from official agencies and foundations is therefore very vulnerable to policy change or expenditure cuts in the major funders.²² A sustainable financing model for IEs and associated synthesis products in LMICs—and particularly in countries with significant dependence on international funding—would require a more diverse financing base among official donors and foundations, and a significant increase in other streams of financing, notably by countries themselves (and, for some synthesis products such as SRs, a greater willingness by international policy communities to co-fund and co-commission such reviews—see below).

Adoption by low- and middle-income countries²³

According to the 3ie IER, the 10 countries where the largest number of IEs had been conducted as of June 2018 were: India (390), China (281), Mexico (247), Kenya (233), Bangladesh (197), South Africa (194), Brazil (193), Uganda (173), Pakistan (105), and Peru (105).

These figures, while indicative of overall activity, do not differentiate between evaluations financed and commissioned by the countries themselves and by other actors, notably the international donor community. In Section 3, we examine in more detail the situation in Mexico, Colombia, Philippines, South Africa, and Uganda.

Mention should also be made of the longstanding commitment of the Tamil Nadu Government to independent evaluation through its Department of Evaluation and Applied Research, which was established as long ago as 1974.²⁴ This has more recently led to a partnership with J-PAL specifically for evaluations of the impact of innovative programmes, strengthening government M&E systems and practices, and enhancing government officials' capacity to generate and use data. To date, this partnership has worked on twelve projects, including five pilot studies, one panel survey, and six randomized evaluations.²⁵

An expanding suite of products

As experience with IEs has grown, various adaptations have taken place in the way such evaluations are designed.

The 'classic' model—and probably still, in quantitative terms, the largest one—is the randomized controlled trial (RCT).²⁶ In this paper, we do not enter into the debates around the virtues and limitations of this approach.²⁷ Instead, we note:

- a) the increasing use of robust experimental and quasi-experimental approaches, which aim to secure comparable rigour in situations where randomization is not feasible;

²² As one example, DFID's share of the financing of 3ie fell from 25 per cent in 2017 to 6 per cent in 2018, as the Development Priorities Window, supported by DFID, approached completion.

²³ In practice it is likely that countries are undertaking evaluations that include the name 'impact' and that may be using quantitative methodologies, but these may well not be using counterfactual approaches, which are at the heart of impact evaluation.

²⁴ Department of Evaluation and Applied Research website: <https://www.dirrear.tn.gov.in/> (accessed 9 September 2019).

²⁵ J-PAL website (accessed 21 July 2019).

²⁶ Martin Ravallion (2018), drawing on 3ie data, states that 60 per cent of all IEs since 2000 have been RCTs.

²⁷ See Ravallion (2018) for a discussion of the views of important researchers such as A. Banerjee, E. Duflo, G. Imbens (pro), N. Cartwright, A. Deaton, J. Heckman, and M. Ravallion (con—or at least cautious about RCTs).

- b) an increasing recognition that evaluability is a significant issue, especially where an IE is commissioned after project start-up;
- c) the positive value of entering into discussions of the utility of IEs at the project design phase (as well as of learning from the increasing databank of knowledge in existing IEs and associated synthesis products). This is often vital for establishing robust baselines, and for assessing the impact of different interventions within the overall project.
- d) the value of formative and process evaluations during the life of the project, facilitating quick response and adaptive programming, rather than leaving impact evaluation to the end of the project;
- e) The development of less expensive and less time-consuming ways of carrying out IEs and synthesis products, while maintaining rigour.

Impact evaluations, capacity development, and the broader evidence system

International agencies suggest that capacity among policymakers to commission and learn from impact evaluation has grown substantially over the past decade; and similar improvements are evident in the ability of ‘southern’ institutions to play a more active part in either leading or participating in IEs (not just as providers of field staff carrying out surveys or accessing data). In both cases, however, countries vary. The various regional Centers for Learning on Evaluation and Results (CLEAR) have played a strong role in developing capacity around evaluation (and M&E more generally), with some of the six centres providing a lot of training in impact evaluation. For example, the South Asia centre is hosted by a regional J-PAL office. Other capacity-building initiatives are being undertaken both by official agencies and by initiatives such as J-PAL, IPA, CEGA, and 3ie.

However, while IEs provide very important evidence to support improved policies and effective delivery of programmes, they should be seen as one important part of a wider evidence system²⁸ that includes other forms of independent evaluation (where these are appropriate), monitoring, results measurement, national statistics, and administrative data.²⁹ In addition there are a range of public good support functions for the evidence system, such as repositories of IEs and SRs (see Box 2).

Box 2: 3ie’s database repositories

The Impact Evaluation Repository (IER) database is an index of all published IEs that meet 3ie’s inclusion criteria. These criteria are that the study must (a) provide a quantitative estimate of the effect of a policy or programme on a development-related outcome; (b) use a rigorous, counterfactual-based method to establish causality; and (c) include transparent reporting about research and analysis methods. As of 2017, there were over 4,000 entries of studies meeting these criteria. The systematic review repository (SRR), like the IER, is open-access and searchable and provides a single point of access for policymakers, programme implementers, and researchers to the largest repository of SRs in international development. The SRR includes user-friendly summaries of high-quality reviews, links to the full texts, and critical appraisal of review methods.

²⁸ Stewart et al. (2019) refer to the evidence ecosystem as ‘A system reflecting the formal and informal linkages and interactions between different actors (and their capacities and resources) involved in the production, translation, and use of evidence’.

²⁹ There are examples of agencies situating IEs within broader M&E systems or investing in data and statistics in parallel, but these currently seem exceptional.

The capacity to consistently manage and supply IEs can be built more sustainably within a functional evidence system. Interventions are typically required across the whole evidence ecosystem, and this needs to be recognized as an investment that will take time and requires sustained local buy-in and support.

The process of evaluations is as important as the quality of the product if a system that is locally owned and sustained is to be achieved. Hence, the way in which IEs are planned, commissioned, and executed affects capacity development and ownership. Our impression, after talking to a wide range of development agencies, is that IEs and syntheses are frequently commissioned in the first instance for learning (and in some cases also for accountability) within the agency itself, with little real buy-in from local stakeholders.³⁰ It would be surprising if such evaluations had much local impact, even if they contributed to the global stock of knowledge.

We do, however, see signs of increasing good practice in some agencies, particularly where local policymakers and researchers are brought together at the project design phase (itself sometimes built on evidence from previous IEs) and where local researchers with adequate experience are given appropriately senior roles in such evaluations, rather than being contracted in to carry out data-related field work alone. This usually happens when evaluations are commissioned by countries, when local research institutions have initiated the idea of undertaking a new evaluation and when local research capacity has grown over time. In fact, as in any other ‘industry’, when evaluation systems have been consolidated within countries, local researchers move upwards in the research ladder and take senior roles, both in the technical evaluation process and in senior positions in public policy.

The case of synthesis products

Single studies are a risky basis for generalizing results,³¹ even though it is possible to assess in a structured way the likelihood of successful application of an approach in other contexts (Bates and Glennerster 2017). There is particular value in looking across rigorous evaluations of particular issues or interventions for an understanding of the underlying issues and the factors shaping the thinking of policy communities.³² And the expanding number of IEs has made useful syntheses more feasible and indeed necessary.

As a result, the expansion of IEs has been complemented by a similar surge in synthesis products, such as systematic reviews, which deliberately scour the universe of rigorous evaluations in order to assess the likely effects of particular interventions in a variety of contexts (White 2019). Table 3 summarizes the numbers of SRs that can be seen in the 3ie repository and in the Cochrane and Campbell libraries³³ when searching for the countries that this paper focuses on. These may not

³⁰ A gradual improvement from a low base in client *involvement* was noted by the World Bank Independent Evaluation Group study of 2012 (Ramirez et al. 2019)—from 22 per cent of IEs initiated pre-2005, to 40 per cent in 2005–2006 and 60 per cent in 2007–2010; but *initiation* of IEs by clients remained very low.

³¹ See for example ‘Be careful what you wish for: cautionary tales on using single studies to inform policymaking’, EJ Jimenez, 3ie News 21 May 2019; available at: <https://www.3ieimpact.org/blogs/be-careful-what-you-wish-cautionary-theses-using-single-studies-inform-policymaking> (accessed 24 January 2020).

³² Thus Banerjee et al. (2016) argue that ‘we need better systems for the production of meta-analyses and review articles and for the creation of expert panels to review the evidence’.

³³ Cochrane (<https://www.cochrane.org/>) has a mission to promote evidence-informed health decision-making by producing high-quality, relevant, accessible systematic reviews and other synthesized research evidence. The Campbell Collaboration (<https://campbellcollaboration.org/>) is an international social science research network that produces high-quality, open, and policy-relevant evidence syntheses, plain language summaries, and policy briefs.

be for individual countries, but for regions (e.g. South Africa is part of Sub-Saharan Africa). This shows that health still dominates, and that the coverage varies greatly between countries.

Table 3: Numbers of systematic reviews by country in different repositories

Repository	Mexico	Colombia	SA	Uganda	Philippines
Health systematic reviews in Cochrane library (nos with search for country name) ³⁴	25	10	78	18	8
Systematic reviews in 3ie repository (nos with search for country name)	13	7	187 (includes SSA)	27	6
Systematic reviews in Campbell library (nos with search for country name)	2	2	6	2	1

Source: the repositories, searched using ‘country’ on 22 July 2019.

Just as the classic IE has often been seen as expensive and slow to deliver results, so have SRs been criticized for high cost, lengthy production time, lack of readability, and often ambiguous conclusions. They remain a very important way of bringing together the findings of studies that meet standards of rigour, but have been complemented by such products as:

- a) *evidence reviews*, often commissioned by individual agencies largely for their own learning, using a variety of protocols;³⁵
- b) *multi-country impact evaluations* of similar issues, with findings brought together in ‘policy insights’ and similar vehicles for dissemination of results;³⁶
- c) *meta-analyses* of relevant groups of impact evaluations.³⁷

In addition, and drawing on the need to investigate the ‘landscape’ of available rigorous evidence, there has been a notable development of evidence maps—sometimes referred to as evidence gap maps (EGMs)—as a tool for setting out where, for a topic area, rigorous published evidence exists. These present a visual overview of existing and ongoing studies or reviews in a sector or sub-sector by intervention type (y axis) and the outcomes measured (x axis).

A review of such evidence maps in international development in 2017 found 73 that met the search criteria, 18 ongoing, and 55 completed, of which 42 had been published in 2015–2016 (Phillips et al. (2017).

Wider sectoral coverage

The evidence from IEs and synthesis products remains predominantly focused on the social sectors,³⁸ as was the ‘When Will We Ever Learn?’ report, but work in new areas has developed

³⁴ Matching ‘countryname’ in Title Abstract Keyword.

³⁵ As an example, DFID commissions evidence reviews, which typically draw on existing SRs, where possible, but are supplemented by more recent studies; and rapid evidence assessments, which provide a more structured and rigorous search and quality assessment of the evidence than a literature review but are not as exhaustive as an SR.

³⁶ For example, J-PAL’s work on micro-finance, based on RCTs in seven countries (J-PAL Policy Insights › Finance › Microcredit: impacts and limitations).

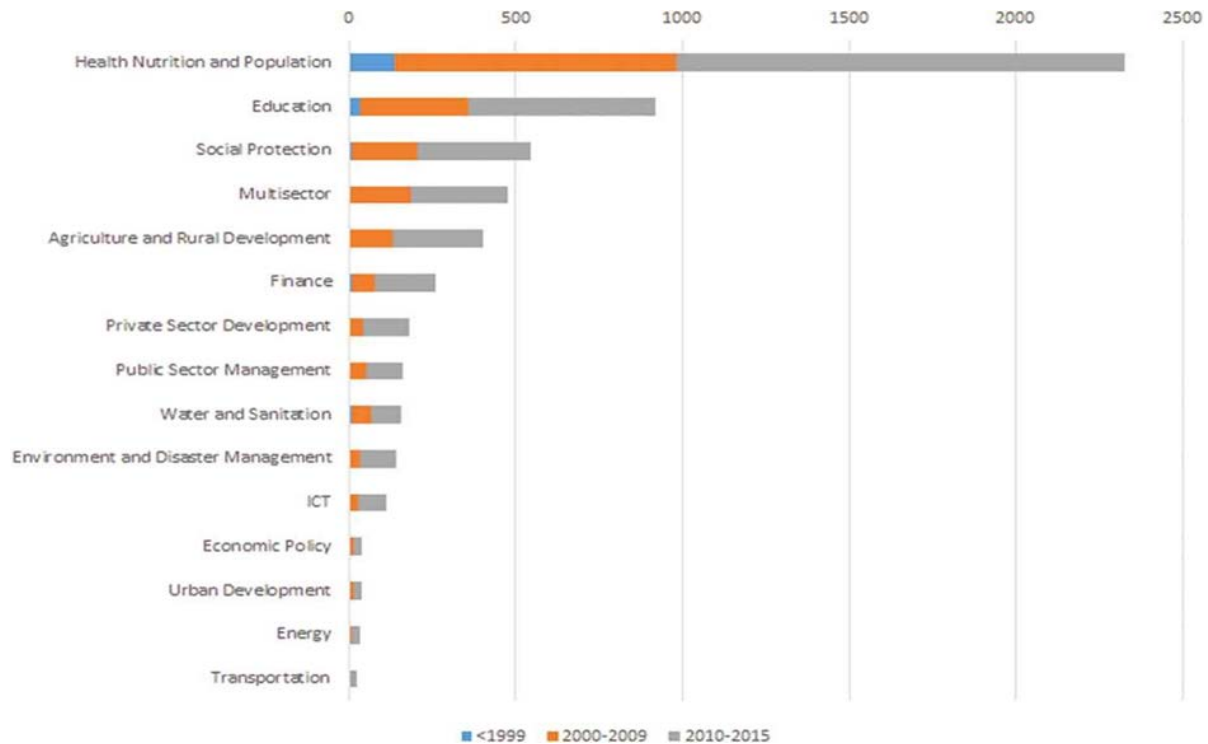
³⁷ For example, the influential study by Fox (2015) on social accountability.

³⁸ For example, in a search conducted on 22 July 2019, 2,595 of the 4,802 IE records fell within the Health, Nutrition and Population categories, and 481 of the 692 SRs. In contrast, only 6 SRs and 44 IEs fell within the category of Economic Policy.

over the past few years, including justice, infrastructure, private sector development, and various aspects of governance (Figure 3).

This expansion of sectoral coverage has demonstrated the scope for using experimental and quasi-experimental designs much more widely than has traditionally been the case.

Figure 3: Impact evaluations by sector and period



Source: Sabet and Brown (2008). Courtesy of 3ie.

2.2 Evidence of use

While the ‘When Will We Ever Learn?’ report put the emphasis on the production of IEs, the large increase in their number poses the issue of whether their findings are themselves having impact. Our interlocutors have varying views on this, some taking a generally positive view, while others emphasize that much depends on local circumstances. One cautioned that ‘The generation of evidence far outstrips the adoption of it. While funders, policymakers, and practitioners have greater access to more evidence than before, use of evidence is by no means systematic or monotonically increasing.’

We noted in Section 1.2 the scarcity of independent studies on the use of the findings of IEs and synthesis products. In this section, we first report on such independent studies as we have been able to find, and then consider self-reported impact by various financiers and providers of IEs.

Existing independent studies

a) World Bank

In 2012, the World Bank’s Independent Evaluation Group examined the relevance, quality, and use of the rising number of the Bank’s IEs (Ramirez et al. 2012).

Under use, it observed that:

1. The use of World Bank IEs to provide evidence of program impact or to inform operational decisions [instrumental use] was ‘modest’. Fewer than half of completed World Bank IEs were mentioned in the project completion documents to demonstrate project impact. [However, this finding should be read with 4 below, which suggests that around half of IEs were used to influence policy dialogue.]
2. The incidence of World Bank IEs being used for decisions to continue, expand, scale down, or cancel the evaluated project or to initiate and refine the design of follow-on projects [also instrumental use] was also sporadic (one-third of cases at most).
3. Some IEs had influenced projects beyond the ones they evaluated. The study particularly notes how the positive IE findings and lessons of the pioneer conditional cash transfer (CCT) program in Mexico (Progresas/Oportunidades) inspired other countries in the region and elsewhere to adopt similar instruments. [This is perhaps symbolic use—demonstrating the benefits of IEs.]
4. IEs were perceived to be useful for World Bank Group staff in policy dialogue with clients and donors [process use]. According to a survey of evaluators and project leaders, around half of World Bank IEs were used to influence policy dialogue with a client government. Examples included:
 - a. The positive impacts of the Familias en Acción CCT in Colombia on consumption, schooling, and health, demonstrated by a set of IEs, helped convince the new government to continue and expand the programs and broaden eligibility.
 - b. IE results of nutrition interventions in Madagascar and Senegal contributed to maintaining political support for the programs.
 - c. IE results helped managers of the Rural Roads Rehabilitation Program in Peru to make a case to the Ministry of Finance for funds to ensure the financial sustainability of the program.
5. Institutional strategies at the World Bank Group had benefited from IEs in areas where there was a large body of evidence, such as education and social protection [symbolic use].

Overall, the direct contribution of World Bank IEs to promoting evaluation capacity and culture had been modest but was now increasing. One-third of completed World Bank IEs were considered by the surveyed team leaders and evaluators to be helping improve World Bank staff/client (or other institution) capacity in the conduct or analysis of IEs. This effect was also increasing, with IEs initiated in 2009–2010 demonstrating significantly higher expectations about building staff/client capacity than IEs initiated earlier.

There were also cases where World Bank IEs had increased the interest of counterparts in strengthening the M&E framework of the evaluated projects and follow-on projects. For example, for 65 per cent of follow-on projects of evaluated interventions, an IE was planned for similar or complementary interventions at the appraisal stage. Country-level case studies also indicated five IEs that had contributed to encouraging governments and project teams to adopt a more evidence-based policy-making culture.

b) Inter-American Development Bank (IDB)

In October 2017, the Office of Evaluation and Oversight of the IDB published a study of the Bank’s production of IEs, their use and impact (Crespo and Azuara Herrera 2017). The report found that, between 2006 and 2016, the Bank had proposed 531 IEs in loan documents and technical cooperation projects. Of these, 94 had been completed by the time of the report, 286 were ongoing (at different stages), and 151 had not proceeded for such reasons as project cancellation, political changes, and problems in design and implementation.

The evaluators considered impact on the Bank, on the clients, and on the wider international community. In brief, their conclusions were:

- The Bank did not systematically use IEs to inform and help define broad approaches at the sector level. However, use of the evidence produced through IEs can improve the theory of change behind IDB operations, promote organizational learning, enrich the stakeholder engagement strategy, and inform the design of ongoing monitoring.
- While the use of IEs was very varied across the Bank, there was a perceptible increase in the use of IE evidence in project design (implementation use). Interviews with staff suggested that an important use of IEs in the country dialogue had been to avoid programmes with limited effectiveness (the example quoted being the ‘One Laptop per Child’ programme, which had been implemented in a few countries before the results of an IE by the Bank showed no impact on children’s achievements. As a result, the programme was cancelled or not approved in at least three countries—Peru, Chile, and Mexico).
- The Progresa/Oportunidades evaluation had served to raise the cost of discontinuing the programme when an opposition government took power, though evidence, gathered over decades, on the dynamics of household allocation of resources had also been relevant. The Progresa IE programme had been key in replicating the model in other countries and in making CCT programmes more palatable and attractive.
- Projects with IEs that were approved between 2009 and 2016 disbursed slightly faster than those without.³⁹
- In a survey of clients, 52 per cent of the interviewees saw impact evaluation primarily as a tool to adjust policy, while 20 per cent identified it as an instrument to support and validate existing policies. There was a wide range of views about the usefulness of the IEs. About 56 per cent of interviewees stated that their (completed) IE had influenced or was expected to influence policymaking. Among interviewees whose IE was ongoing, however, the use or expectation of use in current policy rose to 83 per cent.
- The evaluation found that 58 per cent of IDB’s IEs were unpublished in any form. However, on average each IE was cited 16 times in papers by IDB-affiliated authors.

c) Brazilian municipalities

In June 2019, the study of Brazilian municipalities by Hjort et al. (2019), mentioned earlier, was published. It is an unusual example of applying impact evaluation techniques to address the question of the use and impact of IEs. The paper investigates whether research findings change political leaders’ beliefs and cause policy change.

Collaborating with the National Confederation of Municipalities in Brazil, the authors conducted two experiments with 2,150 municipalities and the mayors who control their policies. In one experiment, they found that mayors and other municipal officials were willing to pay to learn the results of IEs (particularly with large samples), and would update their beliefs when informed of the findings. In the second experiment, they found that informing mayors about research on a simple and effective policy (reminder letters for taxpayers) increased the probability that their

³⁹ This replicated findings in an earlier World Bank study by Legovini et al. (2015).

municipality implemented the policy by 10 percentage points. They see this as direct evidence that providing research information to political leaders can lead to policy change.⁴⁰

d) Assessing how health systems in LMICs learn

Another recent study by Witter et al. (2019) assesses how national health systems in eight LMICs learn from good practice elsewhere. Though by no means limited to the use of evaluations, let alone IEs, its lessons are relevant to much advice from outside, for example:

- The drivers of uptake, or moving ahead with implementation of a policy, are rooted firmly in the local political economy.
- Internal learning is the key to successful policy development over time.
- The case studies illustrated the effective use of annual reviews to assess and improve policy performance, adjustment of policies based on local evidence, using national and international routine data sources for monitoring, use of evidence from operational research, and technical assistance to identify the existing—and possible future—cost structures and affordability of interventions.
- The role of policy evaluation was much more contested. In some settings there was resistance to formally evaluating high-priority national programmes, while in others there were reported tussles over the ownership of the evaluation process. While some countries used evaluations actively as a means of lesson-learning and mid-course corrections, many of the apparently successful policies were never formally evaluated, reflecting the higher stakes and more politicised nature of evaluative processes, compared with continuous learning through observation of a policy's outcomes over time.
- The supply of credible evidence can be facilitated by the development of networks of international and local researchers, producing strong evidence on local policies and building capacity for local analysis.⁴¹

What do the providers and policymakers say?

Major funders, commissioners, clients, and suppliers naturally also assess the use and impact of IEs and synthesis products. From a scrutiny by the authors of published material and websites, and from exchanges with the organizations listed in Appendix B, it is clear that there are many examples of use and impact. Different agencies categorize use and impact in different ways. None of these frameworks follows the categories set out in Section 1.2 exactly, but there is considerable 'matching'. They typically record effects on:

- a) the project, programme, or policy being evaluated (instrumental use);
- b) similar projects, programmes, or policy in the same or other countries (which we see as an example of conceptual use);
- c) global policy thinking about an issue (also conceptual use);
- d) the evidence system, not least in terms of willingness to use experimental and quasi-experimental designs in follow-on or related areas (process use).

⁴⁰ We also examined a Mid Term Review of the SIEF commissioned from the RAND Corporation by DFID (Krapels et al. 2014). Because few studies under DFID financing had as yet been completed, however, this was focused on programme mechanics rather than on use.

⁴¹ See <https://health-policy-systems.biomedcentral.com/articles/10.1186/s12961-018-0410-1>

Examples of *instrumental use* include:

- a) Cases of changes in design, notably where researchers and policymakers/managers work together at the project design stage. For example, a recent DIME assessment⁴² states that 66 per cent of respondents to a survey agreed that DIME's IEs had informed programme/policy design, and 82 per cent agreed that they had helped rationalize existing designs. For its part, 3ie assesses that 30 changes to policy or programme design were observed from a sample of 86 IEs and related projects completed between 2013 and 2018.⁴³
- b) Cases where different treatments are evaluated (DIME, for example, found that 61 per cent of survey respondents agreed that a treatment arm or another element validated by the IE was adopted). 3ie noted 40 cases where studies in the sample quoted above had 'informed discussions of policies or programmes'.
- c) Cases where decisions are needed on whether to scale up a programme or reduce/abandon it. For example, J-PAL researchers found that providing identification cards to beneficiary households improved access to Indonesia's national rice subsidy programme.⁴⁴ The finding informed the Government of Indonesia's decision to scale up social assistance identification cards for a range of programmes targeted at the poorest households across the country. On the other hand, as mentioned above, the IDB reversed an intention to support additional 'One Laptop per Child' programmes after an IE showed no impact on educational attainment. 3ie noted seven cases of scaling-up and two of closing unsuccessful programmes in its survey quoted under a) above.

'Adding up the numbers', as in the statistics given above, is not very satisfactory. For example, an IE may apply to only a small part of a total investment—perhaps to a treatment that affects only a small fraction of potential beneficiaries, or whose positive effect is very modest among the overall factors that affect them. Nevertheless, 'direct' effects of IEs on programmes and projects do appear to be quite frequent, while these remain very inexpensive within the total project value.⁴⁵ The key ratio is the benefit of the evaluation's information to the cost of doing the study. Thus, even a small test evaluation can have huge impact if it affects other projects.⁴⁶

Second, in terms of *conceptual* use, where no direct action has been taken but where people's understanding has been affected, we note many cases where the IE or a synthesis product has influence on other similar projects or programmes, or—and this is likely to be the most important spin-off over time—on wider thinking about the issue being evaluated. Well designed IEs can test general theories about behaviour. Indeed, DFID's Chief Economist stated: 'I strongly believe [that] the way impact evaluation has most impact is by increasing our understanding of underlying

⁴² See *Science for Impact*, DIME's Annual Report for 2019, for details of this and subsequent references to the survey data that underlie these figures.

⁴³ 3ie (March 2019). It should be noted that more than one example of influence is shown for many studies, leading to 147 examples of influence from the 86 products. The 86 products were selected from a database of 234 IEs, 41 SRs, 18 EGMs, and 31 Working Papers as a result of their evident use.

⁴⁴ J-PAL website, Evidence to Policy, accessed 11 September 2019.

⁴⁵ DIME notes (source as for footnote 43) that it has raised US\$180m for its work and 'shaped the design and implementation of more than \$18 billion in development financing'.

⁴⁶ As Bill Savedoff put it to us: 'Think of how many millions of dollars the One Laptop program would have wasted had it not been for the IE'.

issues, not by measuring one-off programmes'.⁴⁷ And, as noted below, other positive effects beyond the project in question are significant.

The 3ie study referred to above shows 27 cases where IEs and associated products informed the design of other programmes, either by direct copying of a new approach validated by an IE within the same country, as with pollution control in India, or by cross-country learning, as with the 'teaching-at-the-right-level' initiative by an Indian NGO now being taken up in 1,800 schools in Zambia.⁴⁸ DIME reports that more than two-thirds of government respondents to its survey said that they used the evidence or data from the IE to guide the design of other projects (80 per cent of which were within their own agencies).

Impact on wider thinking is of course particularly evident in the case of synthesis products. There are some clear cases. The example of cash transfers, both conditional and unconditional (and in humanitarian as well as less extreme situations), is among the best known. The development of WHO guidelines (2016) on adding self-testing for HIV was based on a systematic review which had drawn on IEs managed by 3ie (the only IEs carried out in LMICs), with finance from the Gates Foundation. The move to free distribution of preventive health inputs, such as bed-nets to prevent malaria, is based significantly on a set of studies of the issue by J-PAL. Greater investment in pre-school has been encouraged by many IEs (SIEF, 3ie, etc.), and taken up, for example, in Mozambique. Critiques of 'received wisdom' can be particularly important (for example, J-PAL work on micro-credit or cookstoves). Agencies typically draw on synthesis products, and SRs in particular, in producing their own papers on what the evidence shows, as with an Australian study on education which drew, *inter alia*, on a 3ie SR and an evidence paper financed by a DFID research grant that examined 10 impacts of water, sanitation, and hygiene (WASH) interventions for UNICEF, and which also drew on SRs and published meta-analyses of such reviews.

Whether this progress is as good as it might be is another matter. It is striking that many agencies cite the same examples of such policy impact. International policy communities are themselves quite variable in structure and approach. For obvious reasons, in some areas (notably health) there is a focus on international standards and protocols, whereas in others (e.g. education) there is much less international standard-setting, and benchmarking of outcomes is a more usual tool for assessing potential new approaches.

In almost all cases, the evidence of IEs, and particularly of synthesis products, ought to be central to considerations of good practice, technical guidance, and, where relevant, regulations. It is not, however, evident that this way of thinking is, as yet, entrenched beyond a few sectors. In addition, most synthesis products seem to be commissioned by parties with an interest in the sector or topic rather than by the relevant apex institution (where one exists). It is not obvious that this is the best route to greater use. A greater willingness by international policy communities to demand solid evidence of the likely impact of policies and approaches in their domain would seem more likely to generate a greater sense of collective ownership of ideas validated by IEs and synthesis products.

In addition to the potential 'global' audiences for wider thinking on development issues, there is often—particularly for official development agencies—a further and important internal audience. Thus, good practice in programme x in country y may assist in working up similar programmes elsewhere. Our impression from discussions with such agencies is that this remains a strong

⁴⁷ Rachel Glennerster (personal communication).

⁴⁸ Both quoted in 'Evidence uptake and use from 3ie-supported work: Narrative summaries of examples'. 3ie. Unpublished (personal communication). Ghana is also moving forward with a similar approach (see J-PAL website, under 'Applying Research Insights').

interest (as both the World Bank and the IDB evaluations have shown), but that there are questions about how effective this ‘institutional learning’ is in practice. Impact evaluations are by no means always commissioned by central research or evaluation departments, and indeed a more decentralized approach (evident in DFID, GIZ, or USAID, for example) has its own attractions in terms of engagement with local stakeholders. However, there is a danger that useful learning opportunities will be missed if there is not some way of ensuring that the sector specialists of each agency have a good overview of emerging lessons.⁴⁹

The same is true where a development agency commissions work essentially for improving its own programmes or projects, or indeed in order to improve its accountability to its own oversight bodies. It may be that the inclusion of such studies (if done to acceptable standards) in international databases helps disseminate lessons, but it is not evident that this is usually the case.

The third category, *symbolic use*, is less well documented by the main sources quoted above. The positive side of this is where the existence of an evaluation serves to highlight an issue. It is quite conceivable that the existence of a critical mass of positive IEs of cash transfers has significantly helped the spread of this approach to social programmes. Indeed, the first IE of Progresas⁵⁰ could be seen as a ‘signal’ of the worth of that programme.⁵¹ The negative aspect of symbolic use is most common where IEs are used to justify an existing policy position, sometimes referred to as policy-based evidence.

Finally, *process use*, where the conducting of IEs has had observable consequences in the agencies concerned with implementation and, in some cases, the wider evidence system as a result of the learning that occurs during the evaluation process. Evaluations, and not just IEs, can have significant effects on wider thinking about the use of evidence, even when this may be incidental to their main purpose.⁵²

For example, Goldman et al. (forthcoming), in a study of the National Evaluation System in South Africa (not limited to IEs), note the following ‘unintended benefits’ reported by the departments and provinces that participated in this study: (1) an improved strategic vision as a result of using theories of change; (2) the use of ‘good practice’ in internal research after exposure to external evaluations; (3) an enhanced use of evaluative thinking; and (4) the need to harmonize learning across structures.

The World Bank IEG report (Ramirez et al. 2012) found that 5 of the 19 case study projects with completed IEs ‘contributed’ to more robust M&E strategies, and observed that in 65 per cent of follow-on projects of evaluated World Bank interventions, an IE was planned for either similar or

⁴⁹ The Millennium Challenge Corporation in the US has developed ‘Principles into Practice’ papers, drawing on its IEs, which are open access. DFID has developed some internal syntheses of external evidence of the cost-effectiveness of IEs within different fields: these are not, however, in the public domain.

⁵⁰ The Education, Health and Food Program (Progresas, in Spanish).

⁵¹ Because most IEs in Mexico were conducted initially on ‘good programmes’ (programmes which were seen by governments to have relatively good planning, clear processes, and a positive perception of their value by beneficiaries), IEs became something of a signalling device, as with the Milk Programme.

⁵² Effects may also be visible in commissioning agencies: for example, DFID staff engaged with IEs considered one impact to be an increased focus on evidence (among which robust, high-quality, context-rich evidence is given particular weight) in the drafting of the organization’s Business Cases for new interventions. The Millennium Challenge Corporation (MCC) considers that IEs have resulted in stronger performance evaluations, which can make a credible case of attribution and contribute valuable learning, even where IEs are not feasible (Shreena Patel, personal communication).

complementary interventions. They admit that wider influence is hard to assess, and that IEs are unlikely to be the only contributing factor. They quote as positive examples a project in China where the outcomes of the IE included a significant increase in the government's poverty monitoring and analysis capability and an unusually good database for detailed project monitoring and evaluation; and one in India, where the IE (the first of its type on the topic of HIV/AIDS prevention and other development programmes for migrant workers) helped the client understand the benefits of such evaluation. They also quote the IE of Progres/Oportunidades in Mexico as a major catalyst to the redesign and renewed focus on results of the M&E system of the Ministry of Social Development—a redesign that contributed heavily to the creation of the country's own evaluation council, CONEVAL. Finally, they quote the positive influence that a set of IE activities, some of which were supported by the World Bank, had on Sinergia, Colombia's national results-based management and evaluation system, discussed further in Section 3.2.

The 3ie study on the influence of IEs found 28 examples of IEs 'improving the culture of evaluation evidence use to strengthen the enabling environment'.⁵³ Typical examples were where an IE led to further evaluative work, sometimes involving the original research team (as in pre-school programmes in Colombia), or where skills and interest were built in the implementing agency (as in agricultural extension in Mozambique).

There have also been impacts on thinking and practice about evaluation more generally. As one respondent put it:

The work produced by J-PAL, 3ie, the What Works Centres and Campbell Collaboration have all influenced the rigor, quality standards and norms around impact evaluation and evidence synthesis. While the OECD Development Assistance Committee evaluation criteria were widely recognized as the standard for evaluation, the work of these entities has led to far more nuanced discourse and understanding about impact and what types of evidence can be generated by what types of evaluation methods.⁵⁴

It would be reasonable to conclude that the surge in IEs has helped build a measure of support for these approaches not only in the research community, as already observed, but also in a significant fraction of both implementing and funding agencies, with broader impact on national evidence systems and on the wider practice of evaluation and use of evidence, including that by funding agencies. The extent of this impact is hard to quantify, but it was widely recognized by our interlocutors.

Overall, progress since 2006 is undeniable, though whether it has met initial hopes is assessed further in our conclusions. In the next section we examine what can be learned about the factors that encourage or inhibit use.

2.3 What do international financiers/suppliers say about circumstances favouring or inhibiting impact?

In this section, we first assess our experience and the experiences of major providers of IEs in relation to the factors that encourage or inhibit use and impact in the light of the reviews on research impact more generally by Langer et al. (2016) and Oliver et al. (2014), mentioned in

⁵³ 3ie (March 2019).

⁵⁴ IRC, response to survey.

Section 1.3. We then attempt to explore how far use and impact may be affected by the fact that bilateral and multilateral agencies and foundations fund a large proportion of IEs and synthesis products in LMICs. As we pointed out in Section 1.1, this a distinguishing feature of much of the impact evaluation carried out in relation to these countries (and indeed of other kinds of evaluation as well).

In general, we find that:

- a) Experience underlines the fact that there is not a simple linear relationship between producing evidence and having it used.
- b) There may be different paths for evidence that is of direct relevance to a programme being designed or scaled up (e.g. an IE of a pilot phase) and evaluations of established programmes, where the accretion of knowledge may be more important in assessing its policy effects.⁵⁵
- c) In line with Oliver's findings, timeliness and relevance to the context are very important in determining the likelihood of use. For example, 3ie is planning to require grantees to undertake politically aware 'context analysis', though the evolving political context can be hard to predict. However, findings that support the expansion of a programme may be easier to implement than those that question the impact of existing programmes.⁵⁶ In the latter case, vested interests may be resistant to uncomfortable evidence. However, that is not a reason not to deliver such evidence.
- d) Most institutions that finance development-oriented IEs have given increasing weight to building effective links between researchers and policymakers. This has been one important element in the increasing use of IE techniques at the project design phase (Section 2.1) and in promoting the concept of continued engagement by researchers, both in assessing experiments during a project and in some cases in continuing to advise as the results of IEs are implemented.⁵⁷ As one example among many, J-PAL's Innovation in Government Initiative funds technical assistance to governments to adapt, pilot, and scale evidence-informed innovations that have been previously evaluated with a randomized evaluation (or evaluations) and found to improve the lives of people living in poverty. There is recognition that merely communicating the results of an IE is seldom adequate, as Langer rightly argues (Langer et al. 2016: 10). The media can also be important. The SIEF, for example, has made particular efforts to reach out to journalists. Social media provides another important gateway to influence. We have also been shown several

⁵⁵ Manny Jimenez (personal communication).

⁵⁶ It has been my experience that the political economy of aid will always determine the appetite for and uptake of evidence. Unsurprisingly, evidence that aligns with priors or creates new opportunities (for expansion, greater investments, meeting commitments, credit claiming, etc.) will often be taken up. However, it is also clear that vested interests in maintaining the status quo can limit the uptake of evidence, should that evidence call for change. We have also seen cases where evidence use/uptake happens because an issue is politically expedient for certain groups.⁷ (comment to authors by one user of IEs).

⁵⁷ The OPM comments, for example: 'In our experience the process of producing a product is crucial to the potential for its uptake in future policy. Often this means active engagement with relevant stakeholders from the beginning of a project cycle: (i) at design stage to build awareness of evidence that will be generated, [...] to agree on policy-relevant research questions, and [to] involve stakeholders in the design to build trust in methodologies that are used (so that the focus is on the evidence produced rather than [on] questions about the validity of the evidence); (ii) [...] early in the analysis phase to discuss early results, rather than stakeholders being "surprised" [by] results from a final deliverable [...] [T]he most successful incidences of evidence uptake are when the engagement of researchers does not end at the delivery of a research product, but rather when there is the opportunity for more continued engagement to support the interpretation and application of evidence.'

examples of apparently productive partnerships developed between researchers and local policymakers, but it is not clear that such partnerships are the norm.

- e) ‘Champions’ within governments can be very influential in translating the advice of researchers into actionable policy. Most funders, however, recognize that it is usually inadequate to rely too much on this element, not least since key officials are often posted to other positions. Where champions have been particularly effective,⁵⁸ this seems to be the product not of brief ‘sensitization’, but rather, as emphasised by Langer, of the capability and motivation of these officials (which itself may of course also reflect the political context).
- f) Synthesis products (by no means only, or even particularly, SRs) do appear to be an effective way of changing approaches to intervention where evidence of impact has been inadequate, or where IEs question the effectiveness of the theories of change underlying an intervention. For example, there seems no doubt that successive IEs and the synthesis of their findings have transformed the willingness of many governments to finance cash transfer programmes, and have also encouraged an increasing use of unconditional transfers. Equally, there is evidence that coordinated studies of microcredit by J-PAL and the IPA have forced significant re-thinking of how and when to use this tool.

These findings are in general consistent with the reviews by Langer et al. (2016) and Oliver et al. (2014). We consider, however, that the fact that a high proportion of IEs are ‘donor funded’, ‘donor commissioned’, and ‘donor-country produced’ raises issues that make this field somewhat different from the broader field of research surveyed by these authors. Each of the three aspects has rather different consequences, which we assess briefly below.

Funding by donors (whether bilateral and multilateral official agencies or charitable foundations) carries with it the likelihood that IEs will be supported for reasons valued by the donor. Where these reasons are limited to making a public good available, or to being directly supportive of more effective outcomes in development programmes, there should be a correspondingly direct value to all parties concerned (though there is always the question of whether some direct financial contribution by international policy communities in the first case and local governments or CSOs in the second may strengthen ownership and hence likely use).

However, it is clear that a significant fraction of donor-financed IEs have a prime purpose of accountability⁵⁹ or learning by the development agency concerned, or indeed may be a response by programme staff to incentives to document evidence or to feed a ‘results’ culture.⁶⁰ Of course, accountability and institutional learning at least are valid reasons to commission IEs, but whether, in such circumstances, there is likely to be a close and effective relationship between researchers and local policymakers is doubtful.⁶¹ Discussion with several development agencies also left us with the impression that cross-institutional learning from IEs funded by such agencies is often far

⁵⁸ On the role of champions see, for example, Kingra and Leach (2019), though the study also points out many other factors in the successful evaluation of the impact of unconditional cash transfers on the government.

⁵⁹ See e.g. Crespo and Azuara Herrera (2017).

⁶⁰ See Crespo and Azuara Herrera (2017).

⁶¹ The OPM comments that ‘Different stakeholders may have different priorities and these may not align. This can sometimes be problematic, for example in the case of donor-funded evaluations of government projects. The biggest impact may be derived from specifically tailoring outputs to government policymakers—but their needs do not necessarily align with that of the donor.’

from effective.⁶² We appreciate that very similar arguments can be made about other kinds of evaluation financed by donors.

Commissioning by donors—i.e. national governments or (for synthesis products in particular) international policy communities, as opposed to commissioning by external agencies—would seem to be positive for ownership, even if most funding were to come from donor sources. We have heard some anecdotal evidence to support this, for example in relation to a major evaluation of the Benazir Income Support Programme in Pakistan, commissioned from Oxford Policy Management by the Government of Pakistan. The SIEF is also examining ‘embedding’ a few projects in competent local agencies. DEval commented that they had worked particularly closely with agencies in Philippines on an evaluation of land-use planning in Philippines by establishing a second, local-based reference group in the country (though not to the point of commissioning), and that the extra effort had been productive and had substantially increased ownership for the results and their dissemination.⁶³

We do, of course, accept that commissioning requires skills and effective procurement systems, and it is therefore no surprise that many development-related IEs are commissioned by specialist intermediary organizations, usually funded by official or private donors. We also appreciate the argument that LMIC policymakers draw on lessons from existing research and can act as ‘clients’ and co-creators for IEs that are funded and commissioned by others.⁶⁴ However, it would seem highly desirable for the proportion of IEs commissioned by national governments, southern-based CSOs, and international policy communities to increase over time

As for *donor-country produced* IEs and synthesis products, i.e. those financed and commissioned by donors or by agencies largely financed by donors, an extremely high percentage of principal investigators (PIs) still come from well established academic and other institutions located in developed countries.⁶⁵ It is true that ‘southern’ capacity has increased and may be expected to continue to do so,⁶⁶ thus enabling teams to incorporate southern participants beyond the traditional role of data collectors, but there is a long way to go.

It is not clear whether IEs led by local PIs to similar standards would automatically have a greater chance of achieving use and impact, but one would certainly expect such PIs to have a greater knowledge than their international counterparts of the political, economic, and social context of the issues being evaluated, and be seen as more legitimate interlocutors. Again, while rigorous standards need to be maintained and while, in some cases, there may be value in using external PIs

⁶² See, for example, page 1 of the 2014 report on ‘How DFID Learns’ from the Independent Commission on Aid Impact (<https://icai.independent.gov.uk/report/dfid-learns/>): ‘DFID does not clearly identify how its investment in learning links to its performance and delivering better impact. DFID has the potential to be excellent at organisational learning [...] DFID is not yet, however, managing all the elements that contribute to how it learns as a single, integrated system. [...] Insufficient priority is placed on learning during implementation. The emphasis on results can lead to a bias to the positive. Learning from both success and failure should be systematically encouraged’. We do not consider that DFID is at all unusual in this respect.

⁶³ Joerg Faust (personal communication).

⁶⁴ Cillian Murphy, J-PAL (personal communication).

⁶⁵ For example, the RAND Mid Term Evaluation of the SIEF for DFID (Krapels et al. 2014) states that in the course of 3 consecutive SIEF calls, 55 successful applicants were from North America and 8 from Europe, compared with 3 from Asia, 1 from Latin America, and none from Africa.

⁶⁶ A good example of this progress is that 10 years ago there were no Campbell Collaboration reviews with authors in developing countries, whereas now there are nearly 30 ongoing reviews, many with all-LMIC teams. (Source: Howard White, from inspection of the Campbell library.)

precisely because they are not beholden to local pressures, we feel that an increase in IEs led by local or regional PIs is important in the interests of sustainability, local institutional development, and contributing over time to the local culture of evidence.

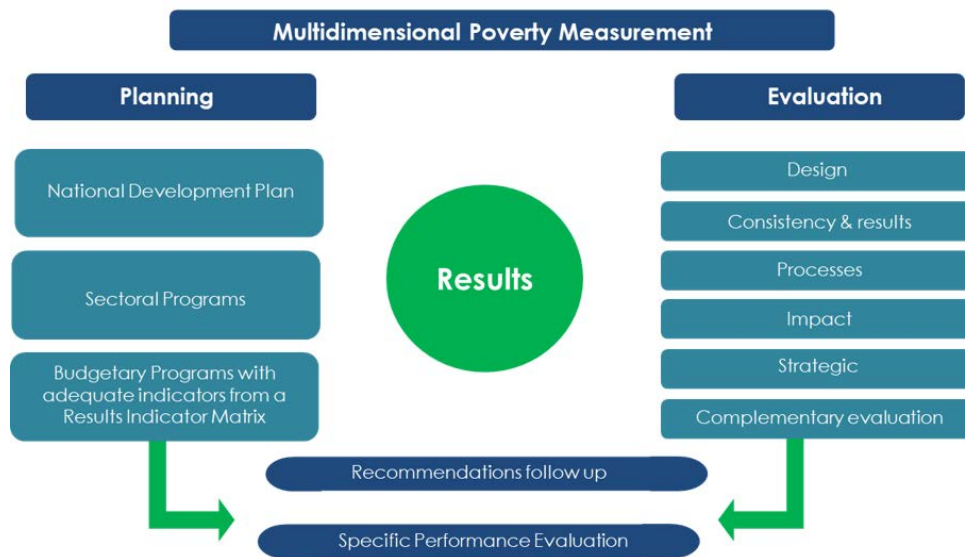
3 Country case studies

3.1 Use and adoption of IEs and syntheses of IEs in the case of CONEVAL in Mexico

Although there were some attempts to conduct IEs during the 1990s, the development of impact evaluation in Mexico was triggered by the experience of the Progresá CCT programme in 1997. In the first evaluation of Progresá,⁶⁷ positive impacts were found on Education, Health and Nutrition, which supported the decision to maintain the programme through the presidential transition of 2000, and to expand its coverage from 250,000 families to over 5 million families.⁶⁸

Progresá is recognized as a breakthrough in the generation and use of evidence for public policy, and is considered as an inspiration for the creation of similar programmes worldwide.⁶⁹ Furthermore, the requirement from the Mexican Congress in 2001 to have external evaluations for every social programme can be attributed to the IE of Progresá, which led to the creation of the Consejo Nacional de Evaluación [National Council for Evaluation] (CONEVAL) in 2006 as an independent institution to evaluate social policy and programmes and, in alliance with the Ministry of Finance, to build the M&E system depicted in Figure 4.

Figure 4: M&E system for social development in Mexico



Source: created by Gonzalo Hernández Licona (co-author) for CONEVAL.⁷⁰

⁶⁷ For further information regarding the IE visit: https://evaluacion.prospera.gob.mx/es/docs/p_docs2000.php

⁶⁸ For a more complete discussion see Levy and Rodríguez (2004).

⁶⁹ It has served as an example replicated in over 60 countries. For a more complete discussion see Banco Mundial (2018).

⁷⁰ Available at: <https://www.coneval.org.mx/InformesPublicaciones/Documents/Memorias/Sistema-de-monitoreo-y-evaluacion-politica.pdf> (accessed 4 January 2020).

As an important element of this M&E system, IEs are incorporated at the most mature stage of the evaluation cycle of interventions to identify whether the intervention is working and to what extent, and to recognize its shortcomings. This facilitates decision-making and accountability. In order to improve the quality of IEs and possible uptake, CONEVAL designed a feasibility analysis outline.⁷¹ This outline allows the institution (CONEVAL) and the programme managers to determine the feasibility of the design, the evaluation characteristics, the methodology, and the type of information to be used.

In Mexico, around 88 IEs have been conducted since 2000, especially in the social sector, including Progresa. These evaluations were financed with public resources as well as by international organizations, as shown in Table 4.

Table 4: Impact evaluations funded in Mexico from different sources

Year	Government funding		International organizations	Total IE
	CONEVAL	Other ministries		
2000		1		1
2001		0	1	1
2002		0		0
2003		0	1	1
2004		4	2	6
2005		2	3	5
2006		2	4	6
2007		3	2	5
2008		2	2	4
2009		2		2
2010		9	3	12
2011		3	6	9
2012	4	2		6
2013		2	2	4
2014		1	7	8
2015	1		2	3
2016	1		4	5
2017	2			2
2018	3	4	1	8
Total	11	37	40	88

Source: authors' arrangement based on information from CONEVAL, Crespo and Azuara Herrera (2017), and 3ie IER.

However, the majority of evaluations conducted by CONEVAL over the past decade are not IEs. The objective is to generate evidence at each stage of an intervention: design process, programme implementation, subsequent monitoring, and impact evaluation. By using different instruments to generate evidence as the programme matures, information is generated at each phase that feeds into the decision-making process. This contributes greatly to the programme's accountability. Since its creation, CONEVAL has coordinated over 2,800 evaluations (Figure 5).

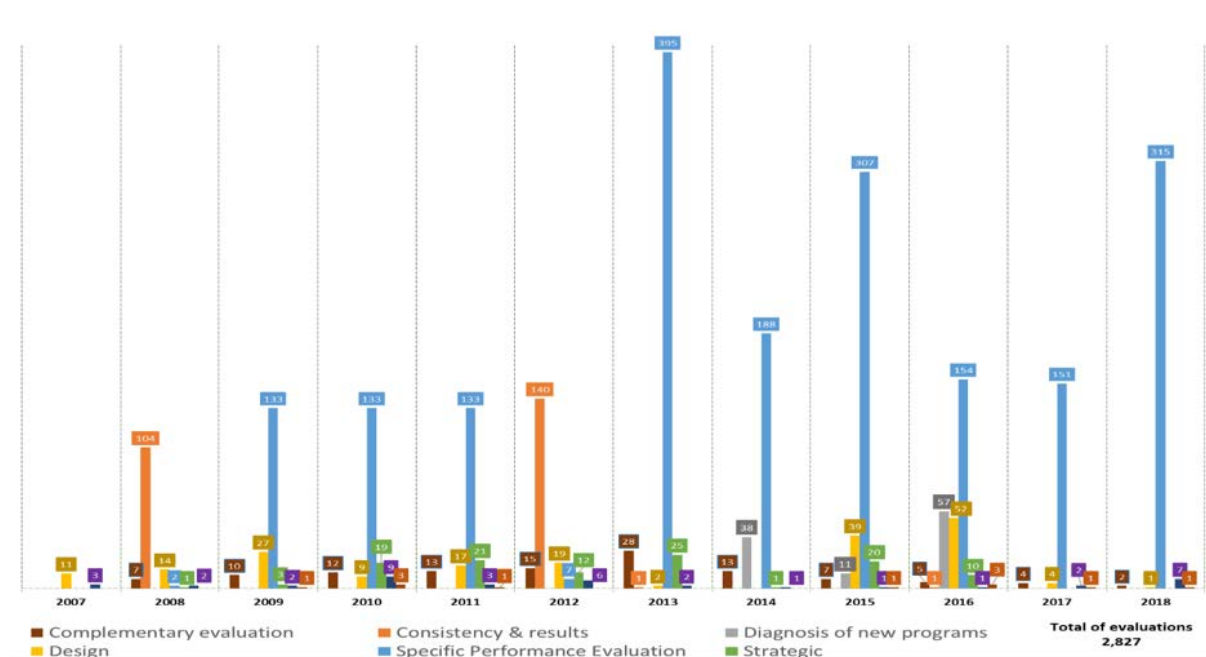
CONEVAL also carries out an annual evaluation of the performance of programmes based on the available information from all types of evaluations. This assessment considers any existing IEs or other evaluations that can provide evidence of the effect of interventions. (It is easier to assess programmes that have an IE.)

⁷¹ For a review of the outline visit:

https://www.coneval.org.mx/Evaluacion/MDE/Paginas/Evaluacion_Impacto.aspx

With this information, programmes can be easily categorized according to their performance. From this, specific recommendations are derived, which are taken up in documents with different objectives in terms of the specific actors these are focused on, with the aim of strengthening decision-making during the different stages of the M&E system. CONEVAL has developed six documents, called ‘Practical Guidelines for Public Policy’, that analyse the existing evidence regarding a specific issue (using evidence maps, SRs, and IEs) based on the country’s characteristics and the diverse interventions, as well as an EGM, in collaboration with 3ie.⁷²

Figure 5: Evaluations coordinated by CONEVAL, 2007–2019



Source: authors' arrangement based on information from CONEVAL.

The use given to evaluation results is also reflected in the Budgetary Consideration Report, intended for Congress, which contains assessments of all social development programmes and specific recommendations for budgeting. This document is presented in a synthetic manner to allow easy and efficient consultation. It is presented to lawmakers and policymakers before the start of budgetary discussions, so that the information is timely for decision-making.⁷³

Evidence of use of evidence for improved policies and programme delivery

The mechanism designed by CONEVAL for following up recommendations from evaluations makes it possible to identify aspects of the improvements in each programme.⁷⁴ The main idea is that programmes have to include an improvement plan, in which they commit to making changes in accordance with the recommendations of the evaluations. The programme-makers prioritize recommendations and engage in the solutions. Future evaluations then track these changes. This

⁷² To access the Practical Guidelines visit:

https://coneval.org.mx/Evaluacion/ESEPS/Paginas/Guias_mejorar_politica_publica.aspx

⁷³ For further information see: <https://www.coneval.org.mx/Evaluacion/IEPSM/Documents/Consideraciones-2019.pdf>

⁷⁴ To consult the Mechanism see:

https://www.coneval.org.mx/rw/resource/coneval/EVALUACIONES/Mecanismo_2011.pdf

follow-up mechanism includes all types of evaluations, which ensures both accountability and the use of the evaluations in improving interventions. CONEVAL periodically records the evaluations' findings and use of recommendations, including quotes from the general public (Table 5).

Table 5: Changes to social programmes based on recommendations from evaluations

Type of improvement	Cycles								
	2011–2012	2012–2013	2013–2014	2014–2015	2015–2016	2016–2017	2017–2018	2018–2019	Total actions
Correct programme activities or processes	163	187	261	79	68	178	232	138	1,306
Modify programme supports	27	12	16	6	19	16	12	11	119
Substantially reorient programme	35	80	80	73	78	109	102	34	591
Add or reallocate programme	3	11	6	1	1	9	8	5	44
Total									2,060

Source: authors' arrangement based on CONEVAL (2018) data.⁷⁵

Findings from impact evaluation have informed public policy debates and led to different degrees of uptake. For example, as a result of the IE of the Milk Programme, the Mexican Government widened the distribution of fortified milk to other programmes, such as the Rural Supply Programme (fortified milk was even sold to private markets, where the higher price compensated for the subsidies of the programme to poorer beneficiaries). The IE of the Food Programme led to a programme redesign that made cash support conditional on attending nutrition talks and receiving nutritional supplements and overall health support. The quasi-experimental IE of the Cement Floor Programme led to the budget being increased.⁷⁶

However, the use of impact evaluation is not isolated. It complements the information generated about each programme by the M&E system, which permits a comprehensive assessment.

Factors favouring or inhibiting use

The emphasis of the Mexican M&E system on the use of evidence has been established by different strategies implemented to encourage the development and use of evaluations. CONEVAL's most important result indicator is the use to which the evaluations are put rather than the number of evaluations produced. Also, the existence of a follow-up system of recommendations makes it possible to systematically monitor the use of evidence, and the adaptation of specific strategies to approach different political and operative actors, although the uptake of SRs is still a challenge in Mexico. Finally, as CONEVAL is responsible for measuring poverty at the national and state level, this measurement generates demand from federal and local governments for good evidence of the effectiveness of programmes in reducing multidimensional poverty.

⁷⁵ The file available in the following link contains this information:

https://coneval.org.mx/quienessomos/ComoNosMedimos/Documents/Usos_informacion_CONEVAL_2007-2017_Excel.zip.

⁷⁶ For further information regarding this IEs visit: https://evaluacion.prospera.gob.mx/es/palprev/p_pal2006.php and www.liconsa.gob.mx/wp-content/uploads/2012/01/impacto_de_la_leche_fortificada_2006.pdf

3.2 Use of IEs in Colombia⁷⁷

The National System of Management and Results Evaluation (Sinergia) has existed in Colombia since 1994. It was created from the National Constitution and Law 154 of 1994, whereby M&E was tied to the process of public policy planning. Sinergia was created to strengthen the process of planning and government management, as part of the role of the Department of National Planning (DNP). In 2004, the three components of Sinergia were established, i.e. monitoring; evaluation; and dissemination of results for accountability.

In the beginning, the evaluation system relied mostly on external resources from international organizations such as the World Bank and the United Nations Development Programme (UNDP). Several government agencies evaluated their biggest programmes at the time with the technical support of Sinergia. Familias en Acción [Families in Action], the Colombian CCT programme, Jóvenes en Acción [Young People in Action], Empleo en Acción [Employment in Action], and Hogares Comunitarios de Bienestar [Wellbeing Community Households] are examples of interventions evaluated using IEs with international funding between 2004 and 2008. Because of the priorities of the supporting organizations, almost every evaluation conducted before 2010 was an IE.

During that period, very few evaluations were funded using Sinergia's own resources, the most significant being the IE of the programme Vivienda de Interés Social Urbana [Urban Social Housing]. Motivated by the international support, government agencies started to devote resources to evaluating their programmes, with Sinergia as a partner providing technical supervision and sometimes administrative supervision.

A change in the evaluation system in Colombia occurred in 2008, when Sinergia began to gain recognition among government agencies and more resources were allocated to the DNP to conduct evaluations itself. Since then, Sinergia has gradually evolved, with the development of a standardized and participatory process for conducting evaluations of different types according to the needs of programmes. To date, 182 evaluations have been conducted by Sinergia, or by other government agencies with the support of Sinergia, 44 of them IEs (Figure 6).

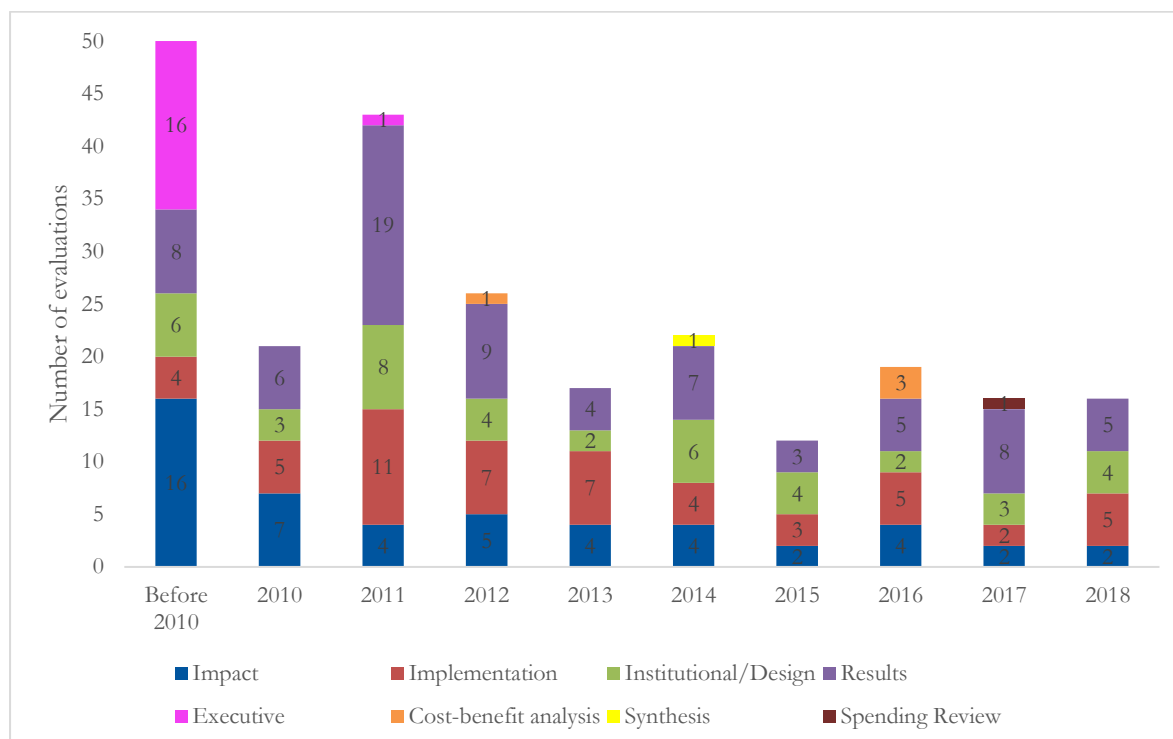
Before the DNP had its own funds for evaluation, 16 IEs were conducted, Familias en Acción (FA) being one of the most evaluated programmes. A solid and independent evaluation and monitoring system was implemented around FA, a comprehensive impact assessment was undertaken, and the World Bank and IDB demanded the implementation of an operational evaluation system referred to as 'spot-checks'. The diagnoses provided by these evaluations led to several modifications of the original programme by the end of 2000.

Through this process, the programme was shown to increase school attendance, mainly among children aged 12 to 17, increase literacy rates (mainly in rural areas), and reduce educational lags. However, it did not lead to better educational outcomes and academic performance. These findings led to a restructuring of the programme's institutional arrangements.

In terms of effects on the labour market, the programme had impacts only in rural areas (especially for rural women). That finding has led to modifications of the programme's focus from big cities towards a dispersed rural population.

⁷⁷ Contributed by Lucas Gómez and Clara Lorena Trujillo.

Figure 6: Number of evaluations supported by Sinergia, by type



Source: Sinergia.

It can be very difficult to identify IEs undertaken by other government agencies, researchers, NGOs, and civil society organizations (CSOs), yet these can be highly influential. For example, two of the best-known IEs of Familias en Acción were undertaken by private researchers, one of them exploring the impacts on the labour market (Quiroga 2008) and the other aimed at the impact of the programme on youth crime (Camacho and Mejía 2013). To address this difficulty, a repository of evaluations was established in 2015, including all good-quality evaluations on public policy issues conducted by government agencies or private organizations. The work is still ongoing (and new challenges have arisen regarding the quality assessment of these evaluations).

More recently, IE methodologies have spread to a wider range of public policy sectors such as retirement funds, health, education, transport, culture, agriculture, and housing. However, given the complexity of these programmes (both operational and political), IEs have been recently supplemented by other types of evaluations aimed at improving their implementation or the institutional arrangement in order to generate better outcomes.

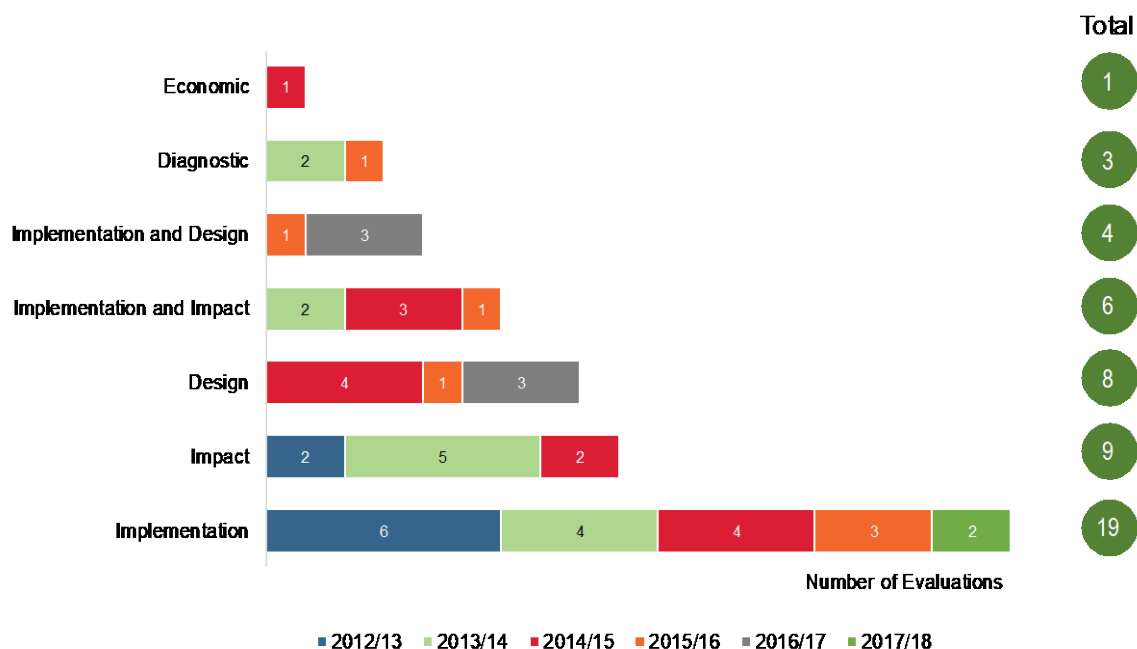
3.3 Adoption of IEs and syntheses of IEs and their use for policy and practice in South Africa

South Africa is a pioneer in the establishment of national evaluation systems (NES) in Africa, along with Uganda and Benin, with its National Evaluation Policy Framework (NEPF) approved in 2011 (Goldman et al. 2018). Sixty-five national evaluations have been completed or are under way, supported by the National Department of Planning, Monitoring and Evaluation (DPME), through the National Evaluation System (Goldman et al. 2019).

It is difficult to know how many IEs have been undertaken in South Africa. One way of assessing this is through the IER of 3ie, which includes 196 evaluations covering South Africa, out of a total of 4,802 evaluations (search of 22 July 2019). These can be split into different evaluation types.

The evaluation of the NES itself, conducted in 2016/17, looked at the evaluations that had been completed at that time and categorized them (Figure 7). This shows that IEs are a small proportion of the total number of evaluations being supported by government.

Figure 7: Types of evaluations supported under the NES



Source: DPME (2018a).

A search in the Cochrane library reveals 78 SRs, showing that SRs are well established in the health sector.⁷⁸ However, this has not translated effectively into other sectors. The number of SRs in the 3ie repository that emerged from a search on South Africa is only 15, and this includes wider regional studies.⁷⁹

An evidence-based policy-making project in the Presidency programme (Programme for Support for Pro-Poor Development) tried to promote SRs and rapid evidence assessments (REAs) but with limited uptake by departments of SRs. Only two REAs were done, and little training in SRs took place. The African Centre for Evidence (ACE) at the University of Johannesburg is the one centre in South Africa specializing in research synthesis outside the health field. It has carried out 14 SRs and 7 REAs.⁸⁰ However, there are now a number of people and universities that have the skills to do experimental and quasi-experimental IEs. These include the University of Cape Town, which is currently collaborating with 3ie and the DPME on a major IE on land restitution; the University of Stellenbosch, which conducted one of the DPME’s IEs on an additional pre-primary year of schooling (Grade R); and the University of the Witwatersrand, which carried out an evaluation on Early Grade Reading.

⁷⁸ South Africa has a strong regional Cochrane Collaboration Centre focusing on research synthesis in the health field (searches undertaken on 22 July 2019).

⁷⁹ Search conducted on 22 July 2019.

⁸⁰ <https://africacentreforevidence.org/2019/02/22/the-africa-centre-for-evidence-growing-the-systematically-collated-knowledge-base-for-evidence-informed-decision-making-in-africa/> (accessed 22 July 2019).

There has also been some innovation in the synthesis field, stimulated by 3ie and adopted by the DPME. The concept of evidence maps/EGMs has been piloted in the DPME, initially in the Human Settlements field, and now a further six evidence maps have been developed by the DPME, as well as 17 by ACE (five of which were a collaboration with the DPME).⁸¹ In addition, the provision of support to policymakers through rapid synthesis has been taken forward in the DPME, working with ACE.

There is also some emerging overlap between these. There has been dissatisfaction with the quality of literature reviews undertaken in NES evaluations, only one using a systematic methodology and drawing from SRs based on IEs, i.e. with strong evidence of what works. The DPME is considering how the quality of literature reviews can be enhanced in this manner, and ran some training at its annual evaluation seminar in 2018.

What evidence is there of influence of IEs on national evaluation/ evidence systems and use?

Some of the IEs undertaken have been key in supporting policymaking and practice. For example:

- The IE of the child support grant was key in supporting the extension of the age at which children were eligible to 18 years, and to counteract the widespread belief that the child support grant encourages teenage pregnancy.⁸²
- IEs of the youth employment tax incentive, along with permission for researchers to access the data, contributed to the decision to expand the scheme.⁸³
- A very convincing IE of the effect of an additional year of pre-primary schooling (Grade R) demonstrated that, while Grade R had been extended widely across the country, it was not contributing to improved learning for poor schools in poorly performing provinces, and the recommendation was to improve quality before scaling up. There have been some instrumental uses: for example, an increase in work on integrated teacher development for Grade R teachers, improvements to the curriculum, greater attention to materials, and development of an integrated ECD plan.⁸⁴

The last example also demonstrates how politics can trump evidence. An electoral pledge of the governing party was to have two additional years of schooling, and, although the advice from the evaluation was to concentrate first on improving quality, the electoral pledge continued for the 2019 elections. This shows that one must be realistic about the use of evidence, and that there is not a simple one-to-one instrumental use of recommendations; nevertheless, there can be an influence on practice.

⁸¹ Harsha Dayal (personal communication, 20 February 2019).

⁸² International document from Department of Social Development on implementation of recommendations of impact evaluation of Child Support Grant.

⁸³ Personal communication (anonymous survey respondent).

⁸⁴ Personal communication (anonymous survey respondent).

Circumstances favouring or inhibiting impact

A book currently in progress, on evidence use in Africa, includes two case studies from South Africa.⁸⁵ These indicate the following conditions for potential government take-up of evaluation results:

- *Political will* to support independent evaluation, information, and data can be observed at the ministerial and senior manager levels and also at the level of project managers.
- *The evidence* being generated is sound.
- There are *champions* in government departments who have been able to convince programme managers that they do not need to worry about negative evaluation findings but rather pay attention to the lessons that emerge for improving implementation.
- *Facilitation role of the DPME* in driving and leading the development of the national evaluation policy framework, developing standards and policy, providing technical advice, and trying to ensure ownership of the evaluation system, both within and outside government.
- The NES places emphasis on *evaluation quality*, with the logic that users are likely to use an evaluation if it is credible and methodologically sound.
- Recommendations are developed *in consultation with stakeholders* through a workshop that allows stakeholders an opportunity to reflect on the recommendations and thereby own and use them.
- *Improvement plans*, and regular progress reports on these, are drawn up, as these are seen as useful steps in institutionalizing the use of evaluation findings.

Some of the barriers to implementation or utilization of the evidence mentioned are:

- *Inadequate institutionalization* of findings, e.g. the findings and recommendations of evaluations are not yet embedded in the annual performance plans of affected departments; nor do the annual reports of these departments reflect progress made in implementing the evaluation results.
- Some of the activities in the improvement plans may require *additional funding*.
- *Complexities* in the structure, resourcing, and scale of programmes.
- The *time needed to undertake evaluations*, especially IEs, which may mean that findings are outdated.

Key limitations on the use of IEs in South Africa have been their cost, the time they take, the availability of data, and whether there is sufficient evidence for effective implementation to justify an IE in the first place. In South Africa, they are generally not designed at the beginning of new programmes and policies to ensure that the data is collected in such a way that impact can actually be seen (randomization, inclusion of a counterfactual, etc.). The evaluation of the NES states:

Going forward, as the system matures, and more departments and provinces buy into the system, it is expected that more impact (as a result of better data through implementation evaluations and improved data management practices broadly) and design evaluations will be conducted (as a result of a better understanding of the need for evaluation throughout the intervention process). A broader range of

⁸⁵ Chapter 5, by Pophiwa et al. (forthcoming).

evaluative information will provide a greater depth of information for decision making and planning (DPME 2018a: 81).

In addition, one of the recommendations is that ‘The role of impact evaluations needs to be strengthened and considered from the beginning of a programme.’ In fact, at present there is some political pressure to look more at impact, and discussions are under way with the DPME as to how this could be done more widely, as well as potentially in the Western Cape Province, a pioneer of the evaluation system at provincial level.

3.4 The contribution of IEs and syntheses of IEs in Uganda⁸⁶

Uganda has a national evaluation system managed from an M&E department in the Office of the Prime Minister (OPM) that has been operating since 2011 (Goldman et al. 2018). A national policy on public-sector M&E was approved in 2013. One aim of the policy is to expand rigorous evaluation of public policy and programmes to ensure that policymakers know what works and what does not. A Government Evaluation Facility (GEF) has been established, run by the OPM, to evaluate public policies and programmes, strengthen accountability for the use of public funds, and improve the design of future policies and investments. Through this facility, some 12 evaluations have been undertaken since 2011, mostly process evaluations. Uganda has collaborated with 3ie to strengthen the implementation of impact and other evaluations. 3ie started by supporting four formative evaluations but has now financed three high-quality IEs and generated evidence that has been used by decision-makers to focus resources on the most promising approaches. 3ie provided guidance and support to produce, synthesize, and quality-assure evidence of what works, for whom, how, why, and at what cost.

The Government of Uganda has also focused on building the capacities of suppliers and users of evaluation and set up an Evaluation Capacity Development Project focused on supporting selected ministries and authorities and CSOs to mainstream professional evaluation in their functions.

Uganda has been a pioneer in Africa in the use of SRs. The Africa Centre for Systematic Reviews and Knowledge Translation at Makerere University specializes in these, and there is also a specialist centre for rapid synthesis, the African Centre for Rapid Evidence Synthesis (ACRES). While initially these focused on health, they have recently been taking a wider approach to include other sectors (Kawooya et al. forthcoming). 3ie and the Campbell Collaboration have provided technical support to build capacity in SRs across sectors. An evidence map conducted by the OPM and Makerere University with the Campbell Collaboration has found that over 500 evaluations have been conducted in Uganda. This provides a useful model that other countries can apply.

Evidence of use

The three IEs that the Ugandan government undertook with the support of 3ie were evaluations of Family Planning, Universal Primary Education, and the Youth Livelihood Programme. The Family Planning evaluation enhanced ongoing debates in the country about the content of the sex education provided to young people in schools. The findings of the Universal Primary Education evaluation informed the design of the new Education and Sports Strategic Plan and the formulation of the Theory of Change to ensure that new programme’s design is informed by particular interventions that will lead to outcomes of interest (conceptual use). The Youth

⁸⁶ Contributed by Abdul Muwanika, Acting Assistant Commissioner for M&E, Uganda.

Livelihood Programme evaluation resulted in a 30 per cent increase in the allocation of funds to beneficiaries (instrumental use).

In terms of use of research synthesis, ACRES supported the Department of Health to develop a food fortification policy, using a systematic review of evidence. This resulted in the policy being adopted nationally.

Circumstances favouring or inhibiting impact

Some of the key factors favouring impact have been:

- Presence of a government knowledge broker stimulating the supply of evaluations and working with sector departments on implementation;
- Openness of IE funders to the initial use of formative evaluations, which can more quickly generate results to feed back into policy;
- Consolidation of efforts by government and donors through the GEF;
- Increased transparency, with evaluations being made public and showcased at various fora such as Uganda Evaluation Week, the National M&E Technical Working Group, and evaluation talk shows organized by the Uganda Evaluation Association, as well as the Parliamentary evaluation platform;
- Commitment by Treasury to use results from evaluations to inform resource allocation;
- For SRs, the presence of a champion in Makerere University.

Some of the inhibiting factors have been:

- Lack of available and accessible quality baseline data;
- Poor timing of impact evaluation studies, which take a long time to provide results;
- Failure to understand what information stakeholders need when designing studies;
- Impact evaluators too concerned with questions of evaluation methodology, so that less attention is paid to the purpose of the evaluation or how it will be used by government;
- A culture that does not accept the kind of criticisms that evaluations inevitably present (Goldman and Olayeye et al. forthcoming);
- Lack of capacity to undertake SRs, and limited awareness outside the health sector.

3.5 Adoption of IEs and syntheses of IEs in Philippines⁸⁷

Since 1992, a total of 66 IEs⁸⁸ have been conducted in Philippines, half of which were carried out in 2011–2017, suggesting increasing demand and growing interest in IEs. This increase was mainly driven by several government initiatives. In 2014, a PhP300 million (US\$6.75m) budget was allotted to the state think-tank, the Philippine Institute for Development Studies, to conduct evaluation studies on key government programmes and projects and IE capacity building for selected government agencies and state universities and colleges. In 2015, 22 process evaluation studies and 6 IE training workshops were conducted.

In 2015, the Australian Government provided AUS\$2.8 million (US\$2.1m) worth of assistance through 3ie for the Policy Window Philippines (PWP), specifically for the conduct of rigorous IEs

⁸⁷ Contributed by Carlos Bernardo O. Abad Santos, Violeta S. Corpus, and William C. Ku.

⁸⁸ 3ie IER: <https://www.3ieimpact.org/evidence-hub/impact-evaluation-repository>.

of programmes on youth employment, sustainable livelihoods, and judicial reforms, and capacity-building workshops on IE for evaluation practitioners in academe, research institutions, and the public sector. PWP is different from other donor-driven initiatives. It is a steering group, chaired by the equivalent of the Planning Minister (the Director General of the National Economic and Development Authority, who is also Secretary of Socioeconomic Planning), that mobilizes the rest of government to identify candidates for IE topics and then decides how to allocate the funds. This builds ownership and capacity in government.

To institutionalize a culture of evaluation in government, the National Evaluation Policy Framework (NEPF) was issued in 2015. This provided for enhanced accountability, informed budget allocations, and better guidance for policy decisions. At the same time, a PhP210 million (US\$4.62m) line-item budget was established for an M&E Fund for the National Economic and Development Authority (NEDA) to support evaluation studies of government programmes and projects and related capacity-building activities.

With these initial initiatives, the evaluation system in Philippines may be classified as in its infancy, since evaluation practice in the public sector remains uncoordinated, but it continues to build on the gains from the various evaluation initiatives. The government is keenly interested in carrying out evidence gap mapping and SRs and enhancing evidence uptake through a proposed PWP Phase II.

Evidence of use for improved policies and programme delivery

Three examples of evidence uptake can be highlighted.

First, the Pantawid Pamilyang Pilipino CCT programme was piloted in 2006. Recommendations from its IE, such as modifying the eligibility criteria, were implemented, which paved the way for the scaling-up of the programme in 2008. Based on the second wave of IE studies, the programme was institutionalized in 2019 through the enactment of a law by Congress.

IE studies of the KALAHI-CIDSS, a community-driven development (CDD) programme, have resulted in improved programme implementation. The third wave of the IE is now assessing the long-term outcomes of the programme, and will determine whether the CDD approach will be institutionalized.

The results of a PWP-supported IE on the Special Program for the Employment of Students resulted in the shift of programme emphasis from improving young people's educational attainment to enhancing employability through the provision of life skills training.

Circumstances favouring or inhibiting impact

The commonly experienced constraint to undertaking impact studies is the lack of availability and accessibility of quality baseline data. An equally important factor is the limited supply of local evaluators. Rigidities in government procurement regulations have meant that there are few qualified bidders willing to tender. To partly address this constraint, NEDA has partnered with the UNDP to commission and conduct thematic (non-IE) evaluations, some of which are expected to produce baseline data for future IEs. This is in effect a means to transfer UNDP knowledge on how to commission and manage evaluations to local M&E specialists.

In addition, a realization from evaluation activities is that it is essential to ensure buy-in from the eventual owners of the study right from the evaluation design, including them in setting the timeline for evaluation activities, and maximizing knowledge-sharing through wider stakeholder

validation and presentation of study outputs. It is also important to emphasize (and clearly differentiate between) impact evaluation, performance monitoring, and other types of evaluation and their intended purposes and outcomes, in the hope of addressing the misconceptions surrounding these types of evaluation, and consequently influence implementers towards programme improvements on the basis of their strengths and weaknesses.

4 Conclusions

4.1 Adoption of IEs

On the basis of our survey, we feel that real progress has been made since 2006 in the adoption of IEs to assess programmes and policies in LMICs. As shown above, this progress has not just been in terms of the number of IEs commissioned, but also in the topics covered, and in the development of a more flexible suite of IE products. There is also some evidence, though mainly anecdotal,⁸⁹ that the insistence of the IE community on rigour has had some effect both in leveraging up the quality of other forms of evaluation and in gaining wider acceptance that ‘before and after’ evaluations with no valid control group tell one very little about the real impact of interventions. In some countries, such as South Africa, Mexico, and Colombia, institutional arrangements have favoured the use of evaluations, including IEs, although more uptake is needed.

There is also perhaps a clearer understanding of where IE techniques can or cannot usefully be applied, or combined with other types of evaluation.

At the same time, some limitations are evident. In the first place, despite the application of IE techniques to new areas, the field remains dominated by medical trials and interventions in the social sectors. Second, even in the health sector, other types of evaluation still account for the bulk of total evaluations, whether by donor agencies or LMIC governments.⁹⁰

Third, despite the increase in willingness of a few LMICs to finance and commission their own IEs, the majority of IEs on policies and programmes in such countries are still financed and commissioned by donor agencies, albeit in some cases with the topics defined by the countries, such as in 3ie’s policy windows. In quite a few cases, the prime objectives of such IEs are domestic accountability and/or learning within the donor agency. We believe that greater local ownership of IEs is highly desirable. While there is much that could not have been achieved without donor finance and commissioning, our sense is that—as with other forms of evaluation—a more balanced pattern of finance and commissioning is needed if IEs are to become a more accepted part of national evidence systems.

Fourth, the vast majority of IEs in LMICs appear to have ‘northern’ principal investigators. Undoubtedly, quality and rigour are essential to IEs, but it is important that IEs should not be perceived as a supply-driven product of a limited number of high-level academic departments in, for the most part, Anglo-Saxon universities, sometimes mediated through specialist consultancy firms. Fortunately, ‘southern’ capacity is increasing, and some programmes have made significant investments in developing this. We take the view that this progress needs to be ramped up very

⁸⁹ See for example the views of DFID and MCC (footnote 51) and of the IRC (footnote 55).

⁹⁰ Raifman et al. (2018).

considerably in the interests of sustainability, local institutional development, and contributing over time to the local culture of evidence.

Fifth, as pointed out in Section 2.1, the financing of IEs depends to a troubling extent on a small body of official agencies and foundations that regard IEs as extremely important products. Major shifts in policy by even a few such agencies could radically reduce the number of IEs being financed.

Finally, while IEs of individual interventions are numerous and often valuable to the programmes concerned, IEs that transform thinking about policies or broad approaches to key issues of development are less evident. The natural tools for such results are more often synthesis products than one-off IEs, and to these we now turn.

4.2 Adoption of synthesis products (building body of evidence)

Systematic reviews and other meta-analyses depend on an adequate underpinning of well-structured IEs, although methodological innovation is now using a more diverse set of sources.⁹¹ The take-off of such products therefore followed the rise in the stock of IEs, and can be regarded as a further wave of the ‘evidence revolution’, as it has been described by Howard White (2019). Such products are increasingly necessary, as the evidence from individual IEs grows.

As with IEs, synthesis products have diversified from full systematic reviews to a more flexible suite of products. We noted examples from international agencies in Section 2.1 and to a lesser extent from countries in Section 3, but many more could be cited. In several cases, synthesis products seek to integrate evidence from quasi-experimental evaluations (e.g. J-PAL’s Policy Insights) or other high-quality research and evaluation evidence.

The need to understand what is now available and where the main gaps in knowledge exist has led in recent years to the burgeoning of evidence maps, pioneered by 3ie but now produced by a variety of institutions and countries. The example of the 500+ evaluations in Uganda cited earlier shows the range of evidence that already exists, which should be mapped and used before new evidence is sought. This should be a priority in all countries.

The popularity of evidence maps shows that there is now a real demand to ‘navigate’ the growing body of IE-based evidence in an efficient manner, as well as to understand the gaps that still exist. The innovation happening also in rapid synthesis shows the demand for synthesis products—but more synthesis is still needed in many sectors and, bearing in mind the expansion in IEs, should be increasingly possible.

4.3 How these products fit within and influence wider evaluation/evidence systems

The concept of an ‘evidence system’ may seem a long way from the often atomized reality of different agencies operating at country level, often in isolation, and sometimes in competition. However, all governments have elements of such a system in place, from national statistical offices to national audit agencies. Often, some part of government, whether a Ministry of Finance or a unit under central government (e.g. a Prime Minister’s office) supervision, has some sort of oversight of at least the main agencies involved. In addition, competent data systems within sectoral bodies and data from national and international surveys make an essential contribution to

⁹¹ For example, 3ie is also including qualitative studies where they provide relevant and well-grounded evidence.

the availability of evidence. Bringing the insights of impact evaluation into the mix is very much work in progress, as the country studies in Section 3 demonstrate.

There is some recognition of the importance of these elements in some of the programmes we have examined (e.g. DIME's work on technology-based data; DFID's support for parts of the evidence system in Uganda described above; CEGA, J-PAL, and 3ie programmes to build local capacity) and there are initiatives, like Twende Mbele in Africa, that look at M&E systems as a whole. However, most IEs are naturally focused on 'the project', extracting what data they can, perhaps with one-off surveys, possibly encouraging further investment in evidence gathering, but making little contribution to a sustainable local architecture. Given the number of IEs now being commissioned, there seems scope for those commissioning IEs and synthesis products to consider how to maximize their positive impact on the development of evidence systems at country level.

We are also seeing the establishment of wide-ranging M&E systems in a number of countries, including Mexico, Colombia, South Africa, and Uganda. In such countries we see in place many elements addressing supply and demand for evaluation and for the use of evidence. These range from policies to standards, competences, custodian organizations for the system, capacity-development processes, and the tabling of evidence in formal organizational processes such as budgeting (Mexico), cabinet (Uganda and South Africa), and national legislatures (e.g. South Africa), as well as training for senior managers in government in the importance of evidence (South Africa, Uganda, Benin).

In countries such as Mexico and Uganda, the role of parliaments has been important not only in the production of evaluations, but also in the construction of the whole system. In the Mexican case, it was parliament that created an independent institution (CONEVAL) to coordinate the evaluation of the social policy; while at the same time, the Law of Budget and Financial Responsibility required a Performance Evaluation System. The role of parliaments could lead to the establishment and consolidation of evaluation systems in many countries.

As can be seen, this is a difficult set of issues to contend with. Constraints are not just organizational and financial, but also political. As with national statistical offices and audit bodies, some degree of independence from political pressures, and some ability to communicate directly to legislatures and the public are highly desirable if truth is to be spoken to power, but, as our own experience has shown, there are limits to the degree of protection that can be sustained in practice, even in systems which on paper appear well protected and robust.

4.4 Evidence of use

In general, we are seeing that the evidence for the use of IEs and SRs is not itself sufficiently rigorous, although we do see some systematic tracking of evidence use—in suppliers such as 3ie, in funders such as the World Bank and IDB, and in countries, such as Mexico, with its follow-up mechanism, and South Africa (DPME 2018b; Goldman et al. forthcoming); and 3ie has been increasing the rigour of its claims by verifying them using contribution tracing (Kingra and Leach 2019).

The use classifications we have applied—instrumental, conceptual, symbolic, and process use—can all be seen in practice, with symbolic the least mentioned (either positive or negative). What is striking is the wide range of uses that must be considered. Some examples are:

- Adjustments to programme activities (63 per cent of social programmes in Mexico; most in South Africa) or major restructuring (29 per cent of programmes in Mexico) or changes in eligibility criteria (in the CCT programme in Philippines);

- Decisions to continue, expand, scale down, or cancel programmes (a survey by DIME showed that 66 per cent of respondents agreed that baselines had informed programme/policy design, and 61 per cent that the treatment arm or another element was adopted;⁹² 3ie noted seven cases of scaling-up and two of closing unsuccessful programmes;⁹³ South Africa noted a youth employment scheme being scaled up (instrumental use));
- Use to inform the design of new programmes (e.g. the evaluation of Universal Primary Education in Uganda);
- Use to inform planning and budgeting processes (e.g. synthesis across evaluations in Mexico; increasing funding for beneficiaries in Uganda's Youth Livelihood Programme);
- Influencing other related programmes, either externally or within the same implementing agency (in addition to well known cases such as Progresá, 3ie reports 27 cases out of 86 studied where IEs influenced other related programmes, and the DIME survey shows that over two-thirds of government respondents agreed that they had observed such influence) (probably conceptual);
- Use for policy dialogue and debate (e.g. milk fortification in Mexico; family planning evaluation in Uganda) (conceptual, symbolic);
- Use to maintain political will (e.g. nutrition in Madagascar; child support grant in South Africa) (symbolic);
- Use to support and validate existing policies (IDB: 20 per cent of cases in the study) (possibly negatively symbolic);
- Use of large bodies of evidence (so, particularly SRs) to inform wider thinking e.g. World Bank strategy; development of WHO guidelines on self-testing for HIV; food fortification policy in Uganda; sector planning around Human Settlements in South Africa) (conceptual);
- Use to justify expanding the use of M&E and IEs (World Bank IEG report; Funza Lushaka evaluation in South Africa) (instrumental, conceptual);
- Improving the culture of evidence use (e.g. 3ie: 28 examples) (conceptual use).

Some of the evidence appears to show a lack of systematic use of IEs to inform sector approaches. However, there are now some studies systematically evaluating the use of evidence, e.g. the study of Brazilian municipalities mentioned in Section 2.2.

Applying the instrumental, conceptual, symbolic, and process use concepts appears most helpful in understanding the psychological process happening in evidence use. The more detailed categories above appear to be more useful for an understanding of what type of use happens.

Overall, the tracking of use is becoming increasingly important. For example, CONEVAL reports that its performance targets are based on use, not generation, and the evaluation of the South African evaluation system indicated that tracking needed to be strengthened. This requires more systematic M&E after the evaluation, but also institutionalization of the post-evaluation process, in relation to both planning and budgeting processes, and individual performance agreements.

As noted in Section 2.2, synthesis products have been linked to significant changes in practice, though, as one respondent put it, 'shifts in international policy communities because of a single

⁹² Legovini et al. (2019: 82, figure 5.1).

⁹³ 3ie (March 2019).

synthesis product are less clear'.⁹⁴ It is striking that a fairly short list of areas where such changes have been seen is mentioned by many organizations involved in commissioning synthesis products and/or IEs. Outside the health sector, where SRs are a longstanding tool for improving medical practice, the list tends to feature welfare (notably cash transfer programmes), education, micro-credit, WASH, and to some extent agriculture.

In general, the various specialist communities that concentrate on key sectors or subsectors would seem to be the most significant recipients of the advice synthesized in these products. This suggests that more frequent commissioning by such communities would in principle be desirable, even if funding comes from one or more donor members.

4.5 Facilitators of use and barriers to use

Earlier we introduced some of the main facilitators and inhibitors of use emerging from the international literature. Table 6 uses these and also draws out other factors in the case study countries and international agencies.

All the cases make the evaluations *available to the public*, and most have some system to ensure the *quality* and reliability of findings, though the raw data themselves are not yet widely available. A key role is played in all cases by (a) government *knowledge broker(s)*, whose role is to link with policymakers and ensure that quality evidence is supplied and that the process is appropriate to ensure *ownership*, and so the likelihood of use of the findings. This may be someone in an LMIC government or a donor programme manager. Apart from the factors suggested by Langer et al. (2016), key is that there is a *system for tracking use*, and most of the country cases show that some system is in place, e.g. an improvement plan, which is tracked. Some suppliers, like 3ie and J-PAL, are also tracking use, as are some funders like DIME, or the IDB. (DFID does not have an agency-wide system for doing this, despite its major investment in evidence generation.)

The role of *champions* is clear—political champions (e.g. ministers) as well as technical champions; both organizations like the DPME or OPM and individuals like the Directors General/Permanent secretaries who head these structures. We also see signs of the impact of *political transitions*. For example, the DPME in South Africa had a stronger mandate in the 2010s, but since the 2019 election the exact role expected of it is not clear, and its Minister in the Presidency no longer carries the title Minister of Planning, M&E. Similarly, the Presidential transition in Mexico has raised questions over the appetite for evidence.

Incentives are key. It is interesting that the fact that CONEVAL also monitors the multidimensional poverty line leads to a demand for interventions that reduce poverty. The DPME uses part-funding as an incentive for evaluations, as does the OPM with its basket fund.

The *institutionalization of an M&E system* is another important facilitator for IE uptake, where actors engage in the decision-making process, production, understanding, and use of evidence (including IEs); Mexico and South Africa are good examples of this.

⁹⁴ IRC, response to survey.

Table 6: Factors influencing use (green=facilitating factor, yellow=mixed, orange=hindering)

Factor ⁹⁵	Examples from:					
	International agencies	Mexico	Colombia	SA ⁹⁶	Uganda	Philippines
Availability and access to research/ improved dissemination	Access to agency-funded or commissioned IEs varies from excellent to poor; 3ie Repository probably the best overall data source	Made public. Repository on CONEVAL ⁹⁷	Made public. Evaluation repository on Sinergia website ⁹⁸	Made public. Evaluation repository on DPME website ⁹⁹	Have evaluation database	Made public. Repository on Socioeconomic Research Portal for Philippines (SERP-P) website. ¹⁰⁰ National Evaluation Portal being developed for NEDA
Clarity/relevance/ reliability of research findings	Standards of analysis vary but high where good quality control exists; relevance beyond concerns of financing agency variable	System designed to ensure quality, with both internal and peer reviews	System designed to ensure quality from the evaluation design stage	System designed to ensure quality	System designed to ensure quality	System designed to ensure quality
Collaboration (between researchers and policymakers)	Emerging good practice by some agencies, but links to local policymakers often weak. PIs predominantly 'northern'.	Agreement and collaboration from the beginning between researchers and policymakers; policymakers participate in recommendations and improvement plans	Sinergia works closely with highly qualified researchers to design and conduct evaluations in a constant dialogue with implementers and policymakers	DPME acts as knowledge broker to make links and ensure process delivers ownership. Procurement system makes partnership difficult. PIs local	OPM acts as knowledge broker to make links and ensure process delivers ownership. Limited supply of PIs for IEs, so often 'northern'.	Technical working groups and/or reference groups of researchers and policymakers created to provide guidance in conduct of evaluation. Validation and stakeholders' workshops promote understanding of evaluation findings and future areas of collaboration

⁹⁵ Initial factors are those shown by Langer et al. (2016) as being correlated with evidence use; others are shown in the cases.

⁹⁶ Some provincial Offices of the Premier are also active and play a similar role in province to the DPME, e.g. Western Cape.

⁹⁷ <https://www.coneval.org.mx/Evaluacion/Paginas/Evaluaciones-y-resultados-de-programas.aspx>

⁹⁸ <https://sinergiapp.dnp.gov.co/#Evaluaciones/EvalFin>

⁹⁹ <https://evaluations.dpme.gov.za/evaluations.aspx>

¹⁰⁰ <https://serp-p.pids.gov.ph/>

	International agencies	Mexico	Colombia	SA	Uganda	Philippines
Timing/ opportunity	General recognition that timing and relevance are important to achieving impact, but in practice agency-specific factors can overrule this	Annual evaluation agenda defines demand; some programmes evaluated each year	Annual evaluation agenda defines priorities based on the National Development Plan	National/Provincial Evaluation Plans define demand. Evaluation time limits demand, especially for IEs	Annual evaluation agenda defines demand.e Evaluation time limits demand, especially for IEs	Donors are primary source of evaluation demand. ¹⁰¹ 5-year National Evaluation Agenda will be developed by government based on agreed evaluation criteria
Policymaker research skills	May exist, but links to local policymakers often weak. Capacity-building programmes in some cases, but effectiveness hard to assess. Agency senior managers have good research skills	Good: knowledge broker in government helps interpret and provides summary reports	Good: knowledge broker in government helps interpret and provides summary reports	Limited: knowledge broker in government helps interpret and provides summary reports. ¹⁰² Fear among programme managers of critical evaluation findings	Limited: knowledge broker in government helps interpret. Fear among programme managers of critical evaluation findings	Studies conducted by PIDS (government think-tank) to assist policymakers to make informed decisions
Costs	Donors (official and private) fund, but funding base remains quite narrow	CONEVAL funds, sometimes with support from agencies such as IDB. IE costs a challenge	National budget allocated both to the National Planning Department and to other government agencies	DPME part-funds, departments fund rest. IE costs a challenge	Use basket fund with multiple donors and government. IE costs a challenge	Donors and some agencies (NEDA and PIDS) fund. Other agencies do not have regular funds for IEs
Other						
Political will	Depends on political drivers, and changes with appetite for evidence. IEs a small part of total evaluation effort for almost all funding agencies	Institutionalization may be strong, due to the political will of Congress and key government officials. Different strategies defined for different actors. Depends on political principle and position taken by President	National Planning Department is engaged and convinced of IEs and their use in public policy	Depends on political principle. Previously had Minister in Presidency.	Minister in Office of Prime Minister	Political support depends on leaders' appreciation of the importance of evaluation ¹⁰³

¹⁰¹ Based on UNICEF (2019).

¹⁰² The introduction of 1/5/25 reports—1-page policy summary, 5-page executive summary, and 25-page main report—has been successful. These are well read, even by ministers.

¹⁰³ UNICEF (2019).

	International agencies	Mexico	Colombia	SA	Uganda	Philippines
Champions	Great variability among agency leadership. New generation of professional staff often trained in IEs and keen to make use of them	Strong champion for evaluation of social policy (CONEVAL) and economic policy (Finance)	Department of National Planning as champion	DPME and some departments and some provinces strong champions	OPM	Apart from NEDA, need for champions in various branches of government ¹⁰⁴
Evidence system	Strong in some agencies as regards their own work, though internal learning variable. Still debates over how to build local evaluation systems, and what priority this should have	Well developed evaluation system, which tracks use	Well developed evaluation system, which tracks use	Well developed evaluation system, which tracks use	Well developed evaluation system	Evaluation system in its infancy
Incentives	For official agencies, typically a mix of 'showing results' as part of accountability, plus improvements to operations and lesson learning within agency. For some (including for institutions specializing in IEs), real commitment to evidence as a public good	Monitoring of CONEVAL's performance is by use, not generation. Widespread use of poverty indicator leads to demand for interventions to address poverty	National Planning Department is the technical arm of the government, and approves the annual budget for different sectors on the basis of their results	National evaluations taken to cabinet. DPME part-funds and provides technical support		No systematic incentive mechanism in place; incentive is foreseen in getting bigger budget for scale-up or replication projects resulting from recommendations of evaluation studies. Award systems recognize implementing units that have developed good practice in project implementation
Data	Only modest evidence of use of production of IEs to improve local data systems	Some IE data available, but more needed	IEs require data collection; all data generated in IEs is made public; some secondary data also made available	Lack of good baseline data for IEs	Lack of good baseline data for IEs	Lack of good baseline data for IEs

Source: authors' arrangement.

¹⁰⁴ UNICEF (2019).

4.6 The implications of a political move away from ‘truth’ in many countries for IEs/syntheses and wider evaluation/evidence systems

As these examples illustrate, the willingness of politicians and top policymakers to take evidence seriously is quite variable, even within a single country. There are particular difficulties when evidence either challenges policies and programmes that have political support or can be portrayed as partial.

On a long view, both the availability and the use of evidence have risen markedly, the health sector usually being a pioneer. In the shorter term, reversals are common, particularly at times of political transition or of closing of the space for debate. All three of us have seen evidence of this in our own careers. Other examples have been brought to our attention in this study, including in countries of major international consequence.

It is impossible to fully inoculate the use of IEs, or a wider system of use of evidence, against political or social constraints. Much depends on the willingness and ability of legislatures and of civil society to challenge executive branches of government. There is also a tendency among new governments to disregard the work of their predecessors. One interesting development in this regard has been the passing of an Act on evidence in the US in 2019, which, however, does not yet appear to be influencing decision-making there.¹⁰⁵

However, the development of capacity to supply and use IEs and associated products should over time enhance the readiness of policymakers and other stakeholders to ask serious questions about the impact of policies and programmes, even if change is slow.

As the example of the study of poverty in the English city of York by Seebohm Rowntree (1901)¹⁰⁶ demonstrated, setting out facts in an accessible manner can have a significant impact on the understanding of an issue over time, even if at the point of publication the issue seems intractable and the proposed solution lacks political support. The answer to challenges to evidence is not to give up on its production. The robust methodology underpinning counterfactual impact evaluations and systematic reviews is particularly valuable in this context.

4.7 Issues for development of IEs/syntheses moving forward

Reflecting on the evidence we have gathered, the progress we have seen, though undeniable, falls short of the ambitions of the ‘When Will We Ever Learn?’ report. We are almost certainly *somewhat* ‘better able to productively use the resources for development, based on an expanded base of evidence about the effectiveness of social development strategies’, but it seems doubtful that the expansion of IEs and associated products has been as transformational as the authors of the report must have hoped when they set the ambition of ‘far better’ use of those resources, on the basis of an expanded investment in impact evaluation and the establishment of what became 3ie.

This is not particularly surprising. Low- and middle-income countries are faced with highly complex problems of sustainable growth, societal change, personal wellbeing, and the development of competent institutions. Many deep-seated issues are not readily amenable to

¹⁰⁵ See https://en.wikipedia.org/wiki/Outline_of_evidence_law_in_the_United_States.

¹⁰⁶ This work proved very influential on the developing political dynamic in the UK, not least in the creation of the welfare state after the Second World War.

standard IE techniques, despite much progress. The pathway to policy influence is seldom straightforward. Interests often inhibit the changes that evidence recommends.

However, the example of the 500+ evaluations that have already been carried out in Uganda shows the range of evidence that exists in many countries, and a priority in most countries should be to map this existing evidence, and to use it to synthesize key themes as needed.

There is a second way in which the CGD report seems not fully to fit the emerging global pattern. Its origin was very much driven by concerns that donor agencies needed to be stricter in applying robust evaluation to programmes it was thought might be the ‘pets’ of, say, individual World Bank programme managers. Although its authors referenced initiatives in Mexico and Chile, and recognized that countries as well as development agencies needed to improve their systems, it underplayed the need for the LMICs themselves to take ownership of the evaluation of the impact of programmes in their countries, whether or not these programmes included donor finance. It paid little attention to how the development of IEs could support improved supply and demand for robust evidence in the countries themselves, and how donors could adapt their approaches to facilitate this. As official development finance declines as a proportion of the gross national income of most LMICs, this perspective needs to change.

The good news is that the increased funding made available over the past 10–15 years has made it possible to experiment with many different ways of bringing researchers and policymakers together in constructive ways, both at country level and in international policy communities, and of engaging not just governments but also legislatures, civil society, and all forms of media. It is time to reflect on what practices seem most likely to promote better evidence systems and better use (always within an uncertain context) of the evidence that IEs and associated syntheses provide. There is much scope for learning between countries, and indeed donor agencies. Donor funding will continue to be important, particularly but not exclusively for lower-income countries, and the priorities for using such funds need careful thought. As noted in Section 2.3, donors may concentrate on questions that are useful for them—which may not be the priority in some countries or adequately prioritize issues of wider concern to the international community. In addition, there appears to be a reduced appetite by donors to fund underlying public goods such as repositories of IEs. As much knowledge is gained by accretion, this would be unfortunate.

We also need to see IEs within a broader spectrum of evaluation, with a wider range of tools available to support policymakers, ranging from evidence maps, rapid evaluations, and rapid synthesis work, to formative/process evaluations and the classic counterfactual impact evaluations and full systematic reviews. The world is more complex than it was, and the needs of policymakers more diverse than a single type of tool can meet. There is a need to develop an evidence agenda to cover the medium- and long-term evidence needs of a sector, at the same time as rapid tools to respond to emerging needs.

The priority in any given country, and the best place to start, will depend on context, the availability of data, the degree to which the M&E system is compliance-based, the political will to learn from failures and undertake evaluation, the capacity in government and among evaluators to undertake evaluations (and IEs specifically), the demand from parliament for evaluations, the pressure from civil society for transparency on performance, and the willingness of donors to support country-led systems. We hope that our survey will encourage greater reflection by all parties on the lessons to be drawn on good practice from the growing body of evidence tools, and from impact evaluations in particular.

References

- Banco Mundial (2018). 'The state of social safety nets 2018' (page 37). Obtained from World Bank: <https://openknowledge.worldbank.org/handle/10986/29115>
- Banerjee, A., E. Duflo, and M. Kremer (2016). 'The influence of randomized controlled trials on development economics research and on development policy'. The State of Economics, the State of The World Conference at the World Bank.
- Bates, M.A., and R. Glennerster (2017). 'The generalizability puzzle'. *Stanford Social Innovation Review*, Summer. Available at: https://ssir.org/articles/entry/the_generalizability_puzzle# (accessed 15 January 2020).
- Camacho, A., and D. Mejía (2013). 'The externalities of conditional cash transfer programs on crime: The case of Familias en Acción in Bogotá'. Documentos CEDE 010552, Universidad de los Andes, Bogotá.
- Cameron, D.B., A. Mishra, and A.N. Brown (2015). 'The growth of impact evaluation for international development: How much have we learned?' *Journal of Development Effectiveness*, 8(1): 1–21. DOI: 10.1080/19439342.2015.1034156
- Center for Global Development (2006). 'When will we ever learn? Improving lives through impact evaluation'. Washington, DC: CGD. Available at: <https://www.cgdev.org/publication/when-will-we-ever-learn-improving-lives-through-impact-evaluation> (accessed 29 January 2020).
- CONEVAL (2018). 'Uso de la información del CONEVAL en la mejora de la política social, 2007-2017.' Mexico City: CONEVAL. Available at: https://coneval.org.mx/quienessomos/ComoNosMedimos/Documents/Usos_informacion_CONEVAL_2007-2017_Excel.zip.
- Crespo, A.R.V., and O. Azuara Herrera (2017). 'IDB's impact evaluations: Production, use and influence'. Washington, DC: Inter-American Bank. Available at: <https://publications.iadb.org/en/idbs-impact-evaluations-production-use-and-influence> (accessed 15 January 2020).
- DPME (2018a). 'Report on the evaluation of the National Evaluation System: Full report (evaluation full report)'. Department of Planning, Monitoring and Evaluation, Pretoria.
- DPME (2018b). 'Report on the evaluation of the National Evaluation System: Summary report (evaluation summary report)'. Department of Planning, Monitoring and Evaluation, Pretoria.
- Fox, J. (2015). 'Social accountability: What does the evidence really say?'. *World Development*, 72 (August): 346–61. <https://doi.org/10.1016/j.worlddev.2015.03.011>
- Goldman, I., et al. (2018). 'The emergence of government evaluation systems in Africa: The case of Benin, Uganda and South Africa'. *African Evaluation Journal*, 6(11). <https://doi.org/10.4102/aej.v6i1.253>
- Goldman, I., et al. (forthcoming). 'Evaluation2—Lessons from 5 years of implementing a national evaluation system in South Africa: Findings from an evaluation'. *African Evaluation Journal*.
- Goldman, I., W. Olaleye, S. Sixolile Ntakumba, M. Makgaba, and C. Waller (forthcoming). 'Mere compliance or learning – M&E culture in the public service of Benin, Uganda and South Africa'. In I. Goldman and M. Pabari (eds), *Using Evidence in Policy and Practice – lessons from Africa*. Abingdon: Routledge.

- Hjort, J., D. Moreira, G. Rao, and J.F. Santini (2019). ‘How research affects policy: Experimental evidence from 2,150 Brazilian municipalities’. NBER Working Paper w25941. Cambridge, MA: National Bureau of Economic Research. <https://doi.org/10.3386/w25941>
- IEG (2012). *World Bank Group Impact Evaluations: Relevance and Effectiveness*. Washington, DC: World Bank.
- Johnson, K., L.O. Greenfield, S.A. Toal, J.A. King, F. Lawrenz, and B. Volkov (2009). ‘Research on evaluation use: A review of the empirical literature from 1986 to 2005’. *American Journal of Evaluation*, 30: 377–410. <https://doi.org/10.1177/1098214009341660>
- Kawooya, I., I. Ddumba, E. Kayongo, R. a Mijumbi-Deve (forthcoming). ‘Rapidly responding to policy queries with evidence: Learning from rapid response services in Uganda’. In I. Goldman and M. Pabari (eds), *Using Evidence in Policy and Practice – lessons from Africa*. Abingdon: Routledge.
- Kingra, J., and B. Leach (2019). ‘How evidence helped address ultra poverty in Malawi’. Evidence Brief Series. New Delhi: 3ie.
- Krapels, J., J. Taylor, M.-L. Henham, B. Baruch, and T. Ling (2015). ‘Strategic impact evaluation fund mid term review: Final report’. Rand Corporation. Available at: https://www.rand.org/pubs/research_reports/RR1137.html (accessed 30 January 2020).
- Langer, L., J. Tripney, and D. Gough (2016). *The Science of Using Science; Researching the Use of Research Evidence in Decision-Making*. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.
- Legovini, A., V. Di Maro, and C. Piza (2015). ‘Impact evaluation helps deliver development projects’. Policy Research Working Papers. Washington, DC: The World Bank. <https://doi.org/10.1596/1813-9450-7157>
- Legovini, A., et al. (2019). *Science for Impact: Better Evidence for Better Decisions – The Dime Experience (English)*. Washington, DC: World Bank Group. Available at: <http://documents.worldbank.org/curated/en/942491550779087507/Science-for-Impact-Better-Evidence-for-Better-Decisions-The-Dime-Experience> (accessed 30 January 2020).
- Levy, S., and E. Rodríguez (2004). ‘El Programa de Educación, Salud y Alimentación, PROGRESA’. In S. Levy (ed.), *Ensayos sobre el desarrollo económico y social de México* (pp. 181–379). Mexico: FCE.OECD (in Spanish).
- OECD (2002). ‘OECD/DAC Glossary of Key Terms in Evaluation and Results-Based Management’. Paris: Organisation for Economic Cooperation and Development (OECD).
- OECD (2006). ‘Outline of principles for impact evaluation’. Paris: Organisation for Economic Cooperation and Development (OECD). Available at: www.oecd.org/dac/evaluation/dcdndep/37671602.pdf.
- OECD (2019). ‘Development aid drops in 2018, especially to neediest countries’. Paris: Organisation for Economic Cooperation and Development (OECD). Available at <https://www.oecd.org/development/development-aid-drops-in-2018-especially-to-neediest-countries.htm> (published 10 April 2019).
- Oliver, K., S. Innvar, T. Lorenc, et al. (2014). ‘A systematic review of barriers to and facilitators of the use of evidence by policymakers’. *BMC Health Services Research*, 14:2. <https://doi.org/10.1186/1472-6963-14-2>
- Patton, M. (1997). *Utilization-focused evaluation: The new century text, 3rd edition*. Thousand Oaks, CA: Sage.

- Phillips, D. et al. (2017). ‘A map of evidence maps relating to sustainable development in low- and middle-income countries: Evidence gap map report’. Pre-Inception Paper. London: CEDIL.
- Pophiwa, N., C. Nuga Deliwe, J. Mathe, and S. Taylor (forthcoming). ‘Using evaluations to inform policy and practice in a government department: The case of the Department of Basic Education in South Africa’. In I. Goldman and M. Pabari (eds), *Using Evidence in Policy and Practice – Lessons from Africa*. Abingdon: Routledge.
- Quiroga, B. (2008). ‘Trabajo infantil en los niños y jóvenes beneficiarios del programa Familias en Acción: una evaluación de impacto’. Archivos de Economía, documento 310, Departamento Nacional de Planeación (in Spanish).
- Raifman, J., F. Lam, J.M. Keller, A. Radunsky, and W. Savedoff (2018). ‘How well are aid agencies evaluating programs? An assessment of the quality of global health evaluations’. *Journal of Development Effectiveness*, 10(2). <https://doi.org/10.1080/19439342.2018.1452779>
- Ramirez, B., E. Javier, and I. Yenice (2012). *World Bank Group Impact Evaluations: Relevance and Effectiveness (English)*. Washington, DC: World Bank. Available at: <http://documents.worldbank.org/curated/en/206961468154467890/World-Bank-Group-impact-evaluations-relevance-and-effectiveness> (accessed 15 January 2020).
- Ravallion, M. (2018). ‘Should the randomistas (continue to) rule?’. CGD Working Paper 492. Washington, DC: Center for Global Development. Available at: <https://www.cgdev.org/publication/should-randomistas-continue-rule> (accessed 30 January 2020).
- Reed, M.S., R. Bryce, and R. Machen (2018). ‘Pathways to policy impact: A new approach for planning and evidencing research impact’. *Evidence & Policy: A Journal of Research, Debate and Practice*, 14(3): 431–58. <https://doi.org/10.1332/174426418X15326967547242>
- Rowntree, B. S. (1901). *Poverty: a study of town life*. London: Macmillan.
- Sabet, S.M., and A.N. Brown (2018). ‘Is impact evaluation still on the rise? The new trends in 2010–2015’. *Journal of Development Effectiveness*, 10(3): 291–304. <https://doi.org/10.1080/19439342.2018.1483414>
- Stewart, R., et al. (2018). ‘Building capacity for evidence-informed decision making: An example from South Africa’. *Evidence & Policy: A Journal of Research, Debate and Practice*, 14(2): 241–58. <https://doi.org/10.1332/174426417X14890741484716>
- Stewart, R., H. Dayal, L. Langer, and C. van Rooyen (2019). ‘The evidence ecosystem in South Africa: Growing resilience and institutionalisation of evidence use’. *Palgrave Commun* 5, 90. <https://doi.org/10.1057/s41599-019-0303-0>
- Storeng, K.T., et al. (2019). ‘Action to protect the independence and integrity of global health research’. *BMJ Global Health*, 2019(4): e001746. <https://doi.org/10.1136/bmjgh-2019-001746>
- UNICEF (2019). ‘Review of national evaluation systems and capacities for evaluating progress towards the Sustainable Development Goals’. New York: UNICEF.
- Weiss, C.H. (1979). ‘The many meanings of research utilization’. *Public Administration Review*, 39(5): 426–31. <https://doi.org/10.2307/3109916>
- White, H. (2019). ‘The twenty-first century experimenting society: The four waves of the evidence revolution’. *Palgrave Commun*, 5(47) <https://doi.org/10.1057/s41599-019-0253-6>
- Witter, S., et al. (2019). ‘What, why and how do health systems learn from one another? Insights from eight low- and middle-income country case studies’. *Health Research Policy and Systems*, 17:9. <https://doi.org/10.1186/s12961-018-0410-1>

Appendix A: Abbreviations

3ie	International Initiative for Impact Evaluation
ACE	African Centre for Evidence, University of Johannesburg
ACRES	African Centre for Rapid Evidence Synthesis (Uganda)
CCT	Conditional Cash Transfers
CDD	Community-driven development
CEGA	Center for Effective Global Action
CGD	Center for Global Development
CIFF	Children's Investment Fund Foundation
CONEVAL	Consejo Nacional de Evaluacion de la Politica de Desarrollo Social [Mexico]
CSO	Civil Society Organization
DAC	Development Assistance Committee [OECD]
DBE	Department of Basic Education (South Africa)
DEval	German Independent Evaluation Agency
DFID	Department for International Development [UK]
DHS	Department of Human Settlements (South Africa)
DIME	Development Impact Evaluation [World Bank]
DJCD	Department of Justice and Cooperative Development (South Africa)
DNP	Department of National Planning (Colombia)
DPME	Department for Planning, Monitoring and Evaluation [South Africa]
ECD	Early Childhood Development
FA	Familias en Acción (Colombia)
GEF	Government Evaluation Facility (Uganda)
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit [Germany]
i2i	Impact Evaluation to Development Impact
IDB	Inter-American Development Bank
IE	Impact Evaluation
IER	Impact Evaluation Repository [3ie]
IPA	Innovations for Poverty Action
IRC	International Rescue Committee
J-PAL	Abdul Latif Jameel Poverty Action Lab
LMICs	Low- and Middle-Income Countries
M&E	Monitoring and Evaluation
NEDA	National Economic and Development Authority (Philippines)
NEPF	National Evaluation Policy Framework
NES	National evaluation system
NGO	Non-Governmental Organization
ODA	Official Development Assistance
OECD	Organisation for Economic Co-operation and Development
OPM	Office of the Prime Minister (Uganda)
PI	Principal Investigator
PWP	Policy Window Philippines
RCT	Randomised Control Trial
REA	Rapid Evidence Assessment
SIEF	Strategic Impact Evaluation Fund [World Bank]
SR	Systematic Review
UNDP	United Nations Development Programme
UNICEF	United Nations Children's Fund
WASH	Water, Sanitation and Hygiene

Appendix B: People consulted

1. Organizations specializing in impact evaluation

Abdul Latif Jameel Poverty Action Lab (J-PAL)

Campbell Collaboration

Center for Effective Global Action (CEGA)

Centre for the Evaluation of Development Policies (EDePo)

Centre of Excellence for Development Impact and Learning (CEDIL)

Innovations for Poverty Action (IPA)

International Initiative for Impact Evaluation (3ie)

Oxford Policy Management (OPM)

Rural Education Action Program (REAP)

2. Multilateral development agencies

Development Bank of Latin America (CAF)

Inter-American Development Bank (IDB)

International Fund for Agricultural Development (IFAD)

Organisation for Economic Cooperation and Development (OECD)

United Nations Children's Emergency Fund (UNICEF)

World Bank: Development Impact Evaluation (DIME)

World Bank: Independent Evaluation Group (IEG)

World Bank: Strategic Impact Evaluation Fund (SIEF)

World Food Programme (WFP)

3. Bilateral development agencies

Department for International Development (DFID)

Federal Ministry for Economic Cooperation (BMZ)

German Agency for International Cooperation (GIZ)

German Institute for Development Evaluation (DEval)

Millennium Challenge Corporation (MCC)

United States Agency for International Development (USAID)

4. Foundations and non-government organizations

Bill and Melinda Gates Foundation (BMGF)

International Rescue Committee (IRC)

Children's Investment Fund Foundation

5. Lower- and middle-income countries

Colombia

Mexico CONEVAL

Philippines NEDA

South Africa DPME

Uganda OPM

Abhijit Banerjee

Cillian Nolan

Howard White

Corey Murray

Orazio Attanasio

Howard White

Nathaniel Goldberg

Emmanuel Jimenez

Marie Gaarder

Beryl Leach

Mark Henstridge

Patrick Ward

Sean O'Leary

Scott Rozelle

Daniel Ortega

Carola Alvarez

Fabrizio Fellini

Sara Savastano

Hans Lundgren

Kerry Albright

Anna Legovini

Alison Evans

Alaka Holla

Silvio Galeano

Rachel Glennerster

Charlotte Watts

Catherine Owens

Alastair Ager

Simon Keogh

Federica Di Battista

Esther Forgan

Clare Winton

Mikaela Zintl

Ricardo Gomez

Franziska Krisch

Joerg Faust

Marion Krämer

Shreena Patel

Tania Alfonso

Ash Shah

Jeannie Annan

Sheree Bennett

Anna Hakobyan

Lucas Gómez

Clara Lorena Trujillo

Thania de la Garza

Janet Zamudio

Paola Vela

Carlos Bernardo Abad Santos

Violeta S. Corpus

William C. Ku

Jabu Mathe

Harsha Dayal

Tim Lubanga

6. Others

UN University, World Institute for Development Economics Research	Tony Addison
Stanford	Ruth Levine
OECD	Hans Lundgren
Center for Global Development	Bill Savedoff
UK Department of Health	Chris Whitty

Appendix C: Checklist for consultees

Respondent details

Name of respondent:

Organisation:

Email:

Telephone:

Date completed:

Date of interview if appropriate:

Prepared to be quoted by name if quote confirmed with them: Yes/No

Role

1. What is the role of your organisation/programme in Impact Evaluation (IE) and/or synthesis products (such as Systematic Reviews (SR) and Evidence Gap Maps (EGM))?

Funder of IEs? Yes/No

Funder of synthesis products? Yes/No

Funder of EGMs? Yes/No

Commissioner¹⁰⁷ of IEs? Yes/No

Commissioner of synthesis products? Yes/No

Commissioner of EGMs? Yes/No

Supplier of IEs? Yes/No

Supplier of synthesis products? Yes/No

Supplier of EGMs Yes/No

Other (e.g. quality control, capacity building)

Please specify

Basic data

2. Where you have indicated 'yes' under funding, how much did you disburse in US\$ equivalent in each of the last 5 years (2014–2018)? Where you have indicated 'yes' under

¹⁰⁷ 'Commissioner' is used to mean that your organisation/programme contracts with the suppliers of IEs and/or synthesis products.

commissioning/supplying, how many IEs and synthesis products did you start in each of those years?

IEs	2014.....	2015.....	2016.....	2017.....	2018.....
Systematic reviews	2014.....	2015.....	2016.....	2017.....	2018.....
EGMs	2014.....	2015.....	2016.....	2017.....	2018.....
Other	2014.....	2015.....	2016.....	2017.....	2018.....

3. If your organisation commissions or supplies IEs and/or synthesis products, who has funded them over the same period? (Indicate where this data may be found if it is in the public domain.)

Overview of adoption of IEs, syntheses of evaluations, and/or EGMs

4. How do you see progress or the lack of it in adoption of individual IEs, synthesis products and/or EGMs for improved policies and program delivery?
5. In which countries do you know that the government (any part of it) has been adopting such products?
6. How have you seen such products fit within wider evaluation and evidence systems? Have you seen or are you aware of such products influencing wider evaluation and evidence systems? Do you have some specific examples you are aware of? If so, can you give links to documentation?
7. Which, if any, international policy communities (ie partnerships that promote or co-ordinate action on specific issues or topics) are you aware of that are adopting IEs or synthesis methods to support their work?

Evidence of use and impact

8. What is your view about progress or the lack of it in use and impact of the findings of individual IEs, synthesis products and/or EGMs in one or more of the following respects:
 - a. Findings and recommendations are at least partially accepted and acted upon?
 - b. The process of carrying out the product and/or knowledge gained by it is influential in future policy?
 - c. The product is significant in raising the profile of the issue being evaluated for policymakers?
 - d. The product is significant conceptually in helping to understand the problem/programme/policy area?
9. Can you quote products in which your organisation was involved that you regard as particularly influential in one or more of these respects?
10. Have you seen or are you aware of synthesis products and/or EGMs influencing international policy communities as defined above? Do you have some specific examples you are aware of? If so, can you give links to documentation?

Circumstances favouring or inhibiting impact?

11. Does your organisation provide specific services related to evidence use?
12. What do you feel are the significant contextual conditions, actions and approaches to evidence uptake and use that favour or inhibit impact:
 - a. In the overall context (political economy, policy, perceived relevance, timing, etc)?
 - b. In the organisational environment or ‘evidence architecture’?
 - c. In the capacities and interests of main actors, including existing champions?
 - d. In use of specific tools (e.g. communications [for example through events, publications, media, social networks]; capacity building; commitment to use; support for implementation)?
13. Can you quote IEs/synthesis products in which your organisation was involved that illustrate the significance of circumstances/approaches to promote evidence use that you have identified?