Miguel Jaramillo Baanante
Lorena Alcázar Valdivia

Impact evaluation and policy decisions: where are we?

A Latin American think tank perspective

September 2015

Working Paper 24

International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding and in producing high-quality evidence of what works, how, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie working papers

These papers focus on current issues, debates and enduring challenges facing development policymakers and practitioners and the impact evaluation and systematic reviews communities. Policy-relevant papers draw on findings from impact evaluations and systematic reviews funded by 3ie, as well as findings from other credible and rigorous evaluations and reviews, to offer insights, new analyses, findings and recommendations. Papers focusing on methods also draw on similar sources to help advance understanding, design and use of rigorous and appropriate evaluations and reviews.

About this working paper

This working paper is a status update on impact evaluation's evolving relationship with public policy, especially in Peru. The authors discuss the inroads impact evaluation has made towards acceptance and influence, what factors have contributed to its advancement and what obstacles remain. They also propose steps needed to enable impact evaluation to contribute more fully to the policymaking process. All of the content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its Board of Commissioners. Any errors and omissions are also the sole responsibility of the authors. The appendixes, available only online, are presented as submitted by the authors. Any comments or queries should be directed to corresponding author Miguel Jarmillo Baanante, mjaramillo@grade.org.pe.

Suggested citation: Baanante, M and Jaramillo, L, 2015. *Impact evaluation and policy decisions: where are we? A Latin American think tank perspective*, *3ie Working Paper 24*. New Delhi: International Initiative for Impact Evaluation (3ie)

3ie Working Paper Series executive editors: Emmanuel Jimenez and Beryl Leach
Managing editor: Annette N Brown
Production manager: Brigid Monaghan
Copy editor: Jaime L Jarvis
Proof reader: Yvette Charboneau
Cover design: John McGill
Printer: Via Interactive
Cover photo: International Center for Tropical Agriculture

Impact evaluation and policy decisions: where are we? A Latin American think-tank perspective

Miguel Jaramillo Baanante
Senior Researcher, Group for the Analysis of Development (GRADE)

Lorena Alcázar Valdivia
Senior Researcher, Group for the Analysis of Development (GRADE)

3ie Working Paper 24
September 2015

International Initiative for Impact Evaluation

Acknowledgements

The authors are grateful to the William and Flora Hewlett Foundation for generous funding for this project. This study was also possible thanks to the active participation of the GRADE researchers and policymakers who shared their impact evaluation experiences with us during long and thoughtful interviews, and to the participants at the workshop on the policy impact of impact evaluation held in Sao Paulo, Brazil. We are also very grateful to the technical team at the Hewlett Foundation, Sarah Lucas and Kristen Stelljes; to Annette Brown and Emmanuel Jimenez of 3ie and an anonymous referee, for insightful comments at different stages of the study; and to participants at the Future of Impact Evaluation conference in Washington DC. Finally, Emilio Legonía and Ruben Neyra provided excellent research assistance.

Executive summary

This paper addresses a question that goes to the heart of impact evaluation research: how much has the field achieved towards its goal of informing public policies and which factors may have been furthering or hindering this goal? Looking primarily at the experience of a Peruvian think tank, the Group for the Analysis of Development (GRADE), we elucidate the importance of factors such as the researcher's contact with programme officers, methodological choices, external validity, institutional factors and public sector capacity, which may all play a role in making impact evaluation results more effective or less effective in informing policy decisions.

Worldwide, impact evaluation research has expanded vigorously in the last two decades. This includes Latin America, where growing interest is evidenced by the number of impact evaluations and the emergence of region-specific research networks working in the field (Cameron, Mishra and Brown 2015; Alzúa, Djebbari and Valdivia 2012). As one of the region's leading research centres, GRADE has contributed to that growth. Its institutional mission is to contribute to Peruvian public policy through rigorous applied research in areas relevant to development. GRADE disseminates the results of the research its associates conduct in order to stimulate and enrich the debate on the design and implementation of public policy. Its researchers also publish in indexed international academic journals; as a group, they lead Peru's economists and institutions in publications available in the Research Papers in Economics online repository. Many of GRADE's researchers go on to hold strategic research and policymaking positions in the public sector.

Between 1998 and 2014, GRADE researchers performed 28 full impact evaluations, in addition to baseline and evaluation design studies that have contributed to impact evaluation. The group's main focus has been the social sector. Social programmes, agriculture and rural development, health and nutrition, and education make up well over half of its impact evaluation studies portfolio. Since 2007 GRADE has conducted an increasing number of impact evaluations, reaching an average of three to four per year. Funding for these studies has come largely from international donor and multilateral institutions (84 per cent), although evidence shows increasing demand from local sources. As demand for impact evaluations has grown and experimental methodologies have gained importance, the group has also seen a transition from non-experimental and quasi-experimental studies to randomised controlled trials since 2008. Before that time, nearly all of GRADE's impact evaluations used quasi-experimental designs.

Regarding the central question of this study, we find that impact evaluations are of increasing interest to the Peruvian public sector, although the greater demand is still for process evaluations. The preference for the process evaluation is associated to the fact that its focus is more directly linked to the everyday tasks and preoccupations of public officials and implementers, making it easier to grasp. Researchers also perceive that public sector officials are incorporating the language

of impact evaluation, albeit mostly among top-level officers. However, use of evidence is still limited and quite heterogeneous across sectors. This evidence suggests that much more can be achieved to support impact evaluations in influencing programme and policy decisions.

By examining a sample of 37 studies conducted by experts from Latin American research centres, we find that only a small portion manage to influence programme or policy decisions, based on researchers' perceptions: only about one in every four studies has managed to have any influence. Moreover, we find that the 'influential' studies share at least two of the following desirable features: intense contact with the implementer from the evaluation design stage, experimental designs, powerful external validity (self-reported by the researcher) and publication of results. However, having these features is not sufficient to achieve influence. Even among impact evaluations characterised by intense contact with programme officers at the design stage, an experimental design and external validity judged strong by the researcher, only half achieved any influence on decisions related to the programme.

Our qualitative analysis, based on interviews, highlights the factors that may be behind these rather modest results in policy influence. The evidence suggests that the dream – of researchers providing the evidence policymakers' need and demand to implement wiser policies – may be unrealistic. Research and policymaking follow different processes, each with its own logic, motivation and timing. Since easy harmony between the two cycles cannot be expected, a question arises as to how to bridge the divide.

Our qualitative data points to several potential strategies, each with its own strengths and weaknesses. The first is for the researcher to work independently, to persuade policymakers of the usefulness of impact evaluation for better policy decision making and thereby foster their collaboration. The second refers to the potential of an evaluation in which the researcher accompanies the intervention design process and applies that information in designing the impact evaluation. A third strategy acknowledges the importance of maintaining a relationship with the policymaker or implementer throughout the impact evaluation process but does not depend on them for the sustainability of the study design – in other words, keeping an appropriate distance.

Although each strategy may be suitable in specific circumstances, a common perception among researchers is that it is not enough to try to persuade public officials about the usefulness of an impact evaluation or to find ingenious ways to get the evaluation done. The keys are to institutionalise learning systems in which impact evaluations play an important role and to establish mechanisms that make their use mandatory, through an entity with authority and suitable capacities to do so. Recent experiences in Latin America suggest that the time is ripe for this type of institutional development.

It is also important to understand that not all public interventions are appropriate for an impact evaluation. Given the considerable cost – in time, funds and expertise –

evaluation questions should be approached more strategically. Collaboration between researchers and policymakers in the task of formulating such questions would also benefit the cause of better policies. Added to these considerations, some evaluations may not be of current interest, although they may be taken up eventually to inform policy. Overall, we find opportunities for policy-relevant impact evaluations in two areas: programme expansion or design of new components for ongoing interventions and underutilised administrative data.

Contents

List of figures and tables

Abbreviations and acronyms

3ie	International Initiative for Impact Evaluation
CAF	Development Bank of Latin America (formerly Andean Development Corporation)
CONEVAL	National Board of Social Development Policy Evaluation
EDEP	Independent Evaluation on Design, Process and Implementation
GRADE	Group for the Analysis of Development
NGO	non-governmental organisation
PRONAA	Programa Nacional de Asistencia Alimentaria
RCT	randomised controlled trial

1. Introduction

Worldwide, the impact evaluation field has seen strong growth in the last decade (Cameron *et al.* 2015). In Latin America and the Caribbean, 70 per cent of the impact evaluations conducted between 1994 and 2011 were produced after 2005 (Alzúa *et al.* 2012). Moreover, new organisations have emerged, focusing their activities on 'the preparation and funding of impact evaluations; and in many countries independent offices have been created for the evaluation and monitoring of public programs' (Bernal 2011). In Latin America, key impact evaluations, such as that of the Progresa/Oportunidades conditional cash transfer programme in Mexico, have contributed to the creation of a global evaluation culture.

Beyond generating new knowledge, a primary motivation for impact evaluation researchers has been to contribute to better public policy implementation. To that end, the field has begun to create a space for itself as the instrument 'to determine if a programme has achieved … the expected results and to explore if there are alternative strategies to achieve such results more effectively' (Gertler 2011). Impact evaluation efforts seek to play a crucial role in the design of public policies, within the now widely accepted framework of evidence-based policymaking. How much the impact evaluation field has achieved towards this goal is the question that guides this study. An essential variable is how an impact evaluation's results inform decisions on the design and implementation of public programmes and policies.[1] Our analysis focuses on this variable, for which current evidence is limited and mostly anecdotal.

This paper presents the results of an approach to the subject from the perspective of the Group for the Analysis of Development (GRADE), a Latin American think tank that, in addition to performing rigorous research, aims to inform public debate (see box 1). The approach uses information collected through surveys and interviews with impact evaluation researchers from GRADE and other Latin America think tanks to address two main questions:

- How much have impact evaluations informed public policies?
- Which factors have favoured or prevented greater influence?

The intent is to help elucidate the importance of factors such as the researcher's contact with programme officers, methodological choices, external validity, institutional factors, supply and demand in the evaluation market and capabilities in the public sector that may play a role in making impact evaluation results more effective in informing policy.

Our results suggest that there is increasing interest in impact evaluation and its potential to inform policies among a fairly heterogeneous set of public sector

[1] By policy decisions, we mean decisions on individual programmes or projects by programme officers or implementers as well as higher-level decisions regarding broader public policy interventions. We are aware that informing decisions at these levels may entail different processes; however, we have chosen to abstract these differences in the interest of identifying general principles.

structures, both between and within countries. In this context, in addition to individual researchers' efforts to design more relevant evaluations, the time seems ripe – particularly in Peru – for institutional developments that give a strong role to learning systems, where impact evaluations hold a crucial place. Some countries, including Mexico and Colombia, have taken substantial steps is this direction. The type of systems to emerge are diverse, from the highly autonomous, such as Mexico's National Board of Social Development Policy Evaluation (CONEVAL), to those anchored in the executive branch of government, such as Colombia's National System of Management and Results Evaluation. Adequacy to the public sector structure is a requisite, as it is credibility.

Box 1: GRADE's Role in Impact Evaluation

GRADE was founded in 1980 as a think tank with the mission of producing independent, rigorous and relevant research to contribute to public policy debates. It has no political affiliation; its activities are funded through research and consultancy services. With 35 years' experience in applied social science research in areas relevant to public development policy, GRADE seeks to disseminate its research not only among academics and experts but also to policymakers, international donors and the general public in order to stimulate and enrich the debate on public policy design and implementation. GRADE's senior researchers all have doctoral degrees from prestigious universities, ensuring that the organisation is up to date and continually innovating in methodological approaches and thematic areas.

GRADE researchers have been involved in impact evaluation projects for the past two decades, beginning with work on youth training and education and public enterprise privatisation programmes. It has now expanded to public programmes in social protection, health, rural development, institutional reform (such as land titling and business licensing) and infrastructure investment. This includes, for instance, the conditional cash transfer programme Juntos ('Together'), looking at outcomes such as household incomes, nutritional status and other welfare indicators of children's welfare. In such studies, GRADE's researchers seek to use the most relevant and rigorous methodologies and to publish their studies in recognised, indexed journals worldwide. In fact, GRADE leads the ranking of institutions and economists in Peru with publications in the Research Papers in Economics online repository (http://repec.org). Many of these articles are based on the results of impact evaluations, and some have been produced in collaboration with international researchers.

A key characteristic of GRADE's research teams – both senior and junior – is their multidisciplinary nature, comprising backgrounds in economics, sociology, anthropology, educational psychology, political science, geography and other fields. For impact evaluation studies, this multidisciplinary nature enables the development of more comprehensive and more insightful studies, thanks to the use of a combination of approaches (theories) and instruments (methods). The use of rigorous, complementary methods makes it possible to account for processes and actors' perceptions and to achieve a better understanding of the interventions and the channels by which impacts occur or do not occur.

GRADE's role is also important in developing new capacities. Junior researchers participate actively at every stage of an impact evaluation and have close contact with public officials. With this experience, many go on to accept positions in the public sector, thereby contributing to the creation of an impact evaluation culture for the Peruvian state.

The remainder of the paper is organised as follows. Section 2 presents a brief review of the literature on the evolution of impact evaluation in the developing world and its contribution to evidence-based policymaking. Section 3 describes our methodological approach. Sections 4 and 5 discuss the state of the field in Latin America and Peru, based on the experience of GRADE and data we collected from a sample of leading researchers based in the region. Section 6 identifies and discusses the main challenges that confront the field, based on interviews with researchers and policymakers. The concluding section points to areas for work oriented towards strengthening the role of impact evaluation in policymaking.

2. Background

According to the Development Assistance Committee of the Organisation for Economic Co-operation and Development, impact evaluation can be defined as the assessment of 'the positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended' (Jones *et al.* 2009). Impact evaluations suitably conducted can provide useful information to researchers, stakeholders and policymakers about whether a development programme accomplishes the desired outcomes.

In the last decade, the number and quality of impact evaluation studies has grown considerably, thanks to the efforts of organisations such as the International Initiative for Impact Evaluation (3ie), the Center for Global Development, the World Bank, the Abdul Latif Jameel Poverty Action Lab and other researchers worldwide. Before 2004 many researchers talked about an 'evaluation gap', referring to the fact that, despite the disbursement of considerable sums on development programmes, few impact evaluation studies had been carried out, with most resources spent on other types of evaluations and monitoring studies oriented to policy needs (Center for Global Development 2006). This gap has been partially filled through the implementation of more and more academically rigorous and policy-relevant impact evaluations. The influence of well designed and effectively implemented impact evaluations, such as the study of Progresa/Oportunidades in Mexico, has helped institutionalise this type of study and increase the perception of impact evaluation as an important tool for policymakers.[2] The interest in cost-effective fund allocation has also supported the propagation of impact evaluations, as we have observed in some of GRADE's experiences, such as the case presented in box 4.

Cameron, Mishra and Brown (2015) find 2,259 published impact evaluation studies conducted in 145 low- and middle-income countries and territories between 1981 and 2012, with annual publication of such studies increasing dramatically after 2008. They find 132 studies published before 2000, mainly in health journals (107, or 81 per cent of the total). After 2000, impact evaluations methodologies and publications begin to appear more frequently in social science journals (p.6).

[2] Two good examples are Behrman and Sengupta (2005) and Behrman, Parker and Todd (2009).

In Latin America, interest in impact evaluation has also grown rapidly. As Alzúa, Djebbari and Valdivia (2012) indicate, two regional research networks are working in the field: the Latin America node of the Partnership for Economic Policy–Policy Impact Evaluation Research Initiative, and the Latin American and Caribbean Economic Association's Impact Evaluation Network. International initiatives such as the Abdul Latif Jameel Poverty Action Lab, Innovation for Poverty Action, Development Impact Evaluation and 3ie also support and implement impact evaluations in the region. However, this type of study is still scarce in Central America and the Caribbean; rigorous impact evaluation is concentrated in Mexico, Colombia, Chile, Brazil, Argentina, Uruguay and Peru. Alzúa, Djebbari and Valdivia (2012) also note that about 70 per cent of the impact evaluation studies conducted in the region between 1995 and 2011 were produced after 2005, indicating a growing trend.

2.1 Impact evaluation prevalence by sector and region

According to 3ie's Impact Evaluation Repository, the worldwide share of impact evaluations published each year has been largest in the health, nutrition and population sector, followed by education, social protection, and agriculture and rural development (Cameron, Mishra and Brown 2015).

Table 1: Impact evaluations published worldwide, 2000 vs. 2012

Sector	2000	2012
Health, nutrition and population	21	233
Education	9	82
Social protection	12	57
Agriculture and rural development	1	48

Impact evaluation studies have also been conducted for a longer period in specific sectors, such as health and nutrition and agriculture and renewable natural resources. The Overseas Development Institute's 2009 report, 'Improving impact evaluation production and use', identifies differences in how impact evaluation studies are used across sectors in the developing world. The authors find more impact evaluations in social development (41 per cent of the data set) and agriculture/natural resources management (23 per cent) than in other sectors. In health and agriculture/natural resources management, the authors identify a growing recognition that experimental impact evaluation can provide robust evidence on relevant questions. They also acknowledge a strong interest in and practice of methodological innovation for impact evaluation in health and social development. In humanitarian aid, infrastructure and results-based aid sectors, however, researchers are more cautious about the applicability of certain methods. The authors also find an increasing recognition of the time it takes to achieve impact, especially in agriculture/natural resources management and infrastructure. Finally, the authors find that commissioning of impact evaluations tends to be supply driven, except in health and social development, where they find growing demand from developing country governments, principally in Latin America.

Alzúa, Djebbari and Valdivia (2012) analyse the state of impact evaluation in Latin America and the Caribbean specifically. Their database comprises 317 impact evaluations conducted in 21 countries over 16 years, from 1995 to 2011. They find that social protection programmes have contributed the largest proportion of studies, partially because most countries in the region have implemented conditional cash transfer programmes, following the successful example of Mexico's Progresa/ Oportunidades programme. There are also multiple studies on each of these programmes, led by the accessibility to impact evaluation databases. In looking at evaluation methodologies, the authors find an increasing trend in the use of randomised controlled trials (RCTs), a methodology that had been used primarily to assess conditional cash transfer programmes.

According to Cameron, Mishra and Brown (2015 p.11),

> Most impact evaluation evidence comes from studies conducted in a single country (96.2%) as opposed to multi-country studies (3.8%) … impact evaluation evidence seems to be concentrated in South Asia (21.9%) and Eastern Africa (19.0%). Meanwhile, South America (14.7%), Central America (10.7%) and Southeast Asia (10.4%) comprise a lower (but still substantial) share of the overall evidence base.

However, since the data set for this study includes only a few non-English studies, both the number and share of South America's impact evaluations may be underestimated. Nonetheless, according to this study the number of impact evaluations published on Latin America – referring to South America, the Caribbean and Central America, including Mexico – has generally been increasing:

- 1980–1990: 4
- 1990–1999: 27
- 2000–2009: 323
- 2010–2012: 241

2.2 Impact evaluation use and policy impact

Theoretically, researchers identify three uses for impact evaluation studies: instrumental (to better design, improve or scale up programmes), legitimising (to justify continued fundraising) and indirect (to increase global knowledge and evidence about what works). However, there is not much known about the use and influence of the impact evaluation studies because of a 'dearth in feedback loops between implementing agencies and evaluators after the evaluation itself has been concluded' (Jones *et al.* 2009, p.225). The reason for this is the lack of incentives for implementing agencies to communicate with researchers on how the evaluation findings subsequently influence the decision-making process, despite the fact that researchers are encouraged to show the impact of their work. Some recent studies have been filling this gap, including Chowdhury, Jenkins and Nandita (2014); Philips *et al.* (2014); Díaz Langou (2013), which focuses on Brazil's Bolsa Familia programme; and Agosto *et al.* (2013), which looks at the Ciudadanía Porteña programme in Buenos Aires.

Jones *et al.* (2009) conclude that legitimisation has been the most common use of impact evaluation findings in practice. This refers to the utilisation of positive findings to justify the actions of the organisation, project or programme to show stakeholders and taxpayers the effective use of funds; in other words, a 'defence of budgets'. This category of use is similar to the 'policy support' category identified by the World Bank in 'Making smart policy: using impact evaluation for policy making', which states that evaluations are 'often used to justify continued funding for a program or to ensure political support for a new or expanded program' (2009 p.10). Nevertheless, according to some researchers, there are some exceptions where well designed and conducted evaluations achieve sound policy influence, such as the impact evaluation of Mexico's conditional cash transfer programme, Progresa/Oportunidades. This study was very relevant for the institutionalisation of impact evaluation in Mexico, contributing, for example, to the creation in CONEVAL in 2006.

At a January 2008 World Bank conference, discussion centred around whether 12 case studies (including Progresa/Oportunidades) had been influential in policy development and why. Participants found that impact evaluations had been influential in helping design better programmes, in legitimising programmes and interventions and in creating a culture of impact evaluation. They also identified factors influencing the use of the impact evaluation findings, such as 'timing and focus on priority stakeholder issues', 'effective dissemination' and 'clear and well communicated messages'. On the other hand, InterAction's fourth 'Impact Evaluation Note', which explores the state of the use of impact evaluation results, suggests several ways to improve the influence of impact evaluation findings: trying to anticipate and plan, at an early stage, how the impact evaluation may be used; frequent communication between researchers and stakeholders from the outset; and creating and enforcing an evaluation culture in the public sector (Bonbright 2012).

In a discussion of the policy influence of impact evaluation, a relevant issue is the increasing demand of more rigorous methodologies, particularly the use of randomised experiments. This trend has been supported by international initiatives such as the Abdul Latif Jameel Poverty Action Lab, the World Bank's Development Impact Evaluation initiative, the Strategic Impact Evaluation Fund and, more recently, 3ie. A typical critique of experimental methodologies that limits impact evaluations' potential policy influence is the 'black box' critique, which refers to the experimental impact evaluation's difficulty in revealing causal pathways from intervention to effects. In other words, impact evaluations do not explain much about how and why an intervention works – or why it does not.

Another problem with the exclusive use of experimental methodology – as well as in some cases of non-experimental methodologies – is its weakness at demonstrating the external validity of results, which limits their usefulness for policy because of policymakers' interest in replicating interventions in other regions or scaling them up. Given the limitations of experimental and quasi-experimental methodologies, especially in relation to the 'black box', in recent years researchers have been endorsing a combination of qualitative and quantitative methods. The combined use of experimental and qualitative methods has made it more feasible to overcome methodological limitations and produce more useful findings for policymakers. We further analyse these and other factors that may favour or impede the use of impact evaluation following section 3, which describes our methodology.

3. Methodological approach

Our analysis uses qualitative and quantitative information in a complementary manner, focusing on the experience of researchers at GRADE and similar research centres from Latin America. We first generated a database of information on GRADE's participation in impact evaluation research studies from the last two decades; we present these data in section 4. Since GRADE's work is funded directly from the studies it conducts, without endowment or contribution from the public sector, we are able to identify some general characteristics of impact evaluation demand in Peru, including the source of the demand, which sectors were more active, the most common methodologies, the reasons behind methodological choices (technical, data-driven or political) and what use was made of the results.

We then gathered survey information from 19 impact evaluation specialists from research centres in 9 countries during a workshop on impact evaluation and policy in Sao Paulo, Brazil, before the annual Latin American and Caribbean Economic Association meeting in November 2014 (see appendix A online).[3] We administered a survey on impact evaluations from the past 10 years, collecting information on the researchers' areas of experience, the sources of demand for impact evaluations, methodological approaches and other factors.

During the workshop we discussed the factors at play and the challenges and lessons learned from impact evaluations carried out by each organisation; the usefulness of the research teams' questions and the most relevant obstacles to informing the policy process and achieving policy impact; the political-institutional context and characteristics of the research teams' incidence strategy; and the challenges to increasing the number of impact evaluations in the region. We present our analysis of these data in section 5.

Building on the information from our analysis of the survey data and our discussions at the workshop, we then prepared guides for in-depth interviews with GRADE researchers and policymakers about their participation in impact evaluations. The questionnaire focused on the impact evaluation process, from how the demand originated to how (and whether) the results were used. Topics included methodology, strengths and limitations of impact evaluation, dissemination and uses of results. We interviewed eight of GRADE's senior researchers, whose experience ranged from social programmes to agriculture, and five policymakers, among them the minister of education, a vice minister at the Ministry of Education, former ministers from the Ministry of Health and the Ministry for Development and Social Inclusion, and a national director at the Ministry of Economy and Finance. We present these findings in section 6.

---

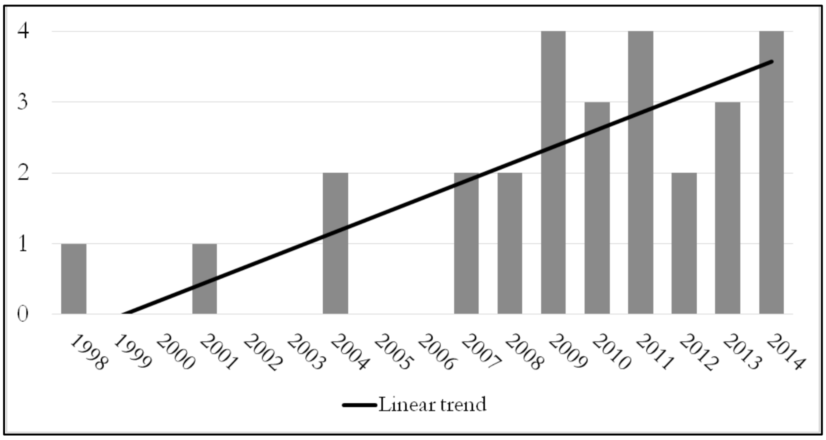
[3] On average, the researchers had each performed 5 impact evaluations in the past 15 years.

4. Impact evaluation trends in Peru

The figures in this section are based on data from the 28 impact evaluation studies GRADE researchers completed between 1998 and 2014.[4] In addition to providing evidence of GRADE's work in this area, they may be good indicators of general trends in the demand for impact evaluation in Peru. This is not only because of GRADE's significant contribution to the field, but also because the organisation does fund its research projects in the market through grants and contracts.

As figure 1 illustrates, beginning in 2007 there is an increasing trend in the number of impact evaluations, reaching an average of three to four per year.

Figure 1: Number of impact evaluations started, by year

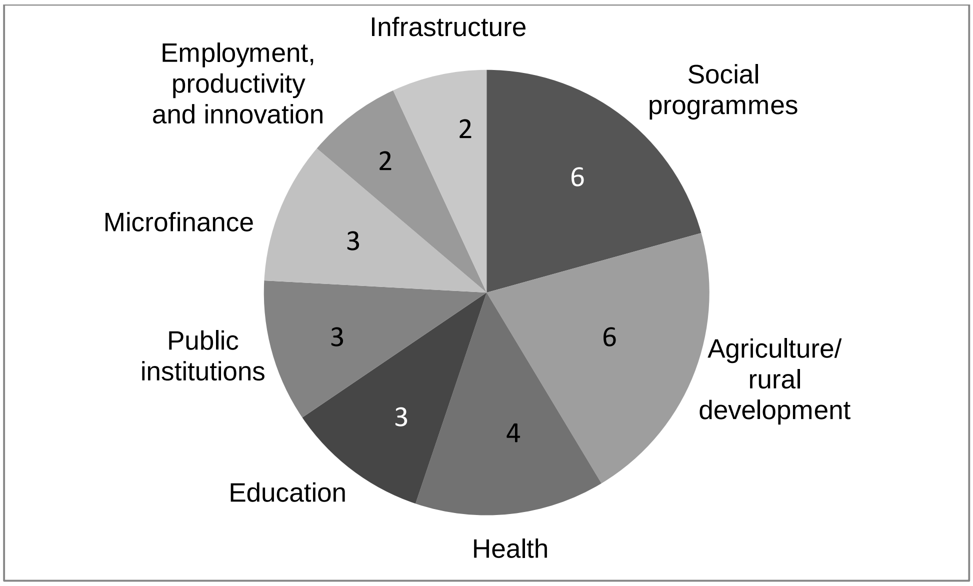


Note: From the GRADE project database.

Impact evaluations have mainly focused on the social sector, which comprises 65 per cent of the studies – social programmes (six studies), agriculture/rural development (six), health (four) and education (three) – followed by public institutions (three studies), microfinance (three), employment, productivity and innovation (two), and infrastructure (two), as illustrated in figure 2.

[4] This sample includes only full studies, not studies that were part of an impact evaluation but did not include the whole process (for example, baseline or evaluation design).

Figure 2: Number of impact evaluations by sector



Note: From the GRADE projects database.

The higher demand for impact evaluation of social and rural development programmes could have two complementary reasons. First, social programmes (excluding traditional social security programmes associated with formal employment) constitute a relatively recent public policy area that has, therefore, a higher demand for justification, whereas activities such as infrastructure are traditional government responsibilities that tend to produce public goods (such as roads and hospitals), so demand for their justification is more limited. Second, social investment is more associated with international donors and multilateral banks, which still represent the main source of demand for impact evaluation.

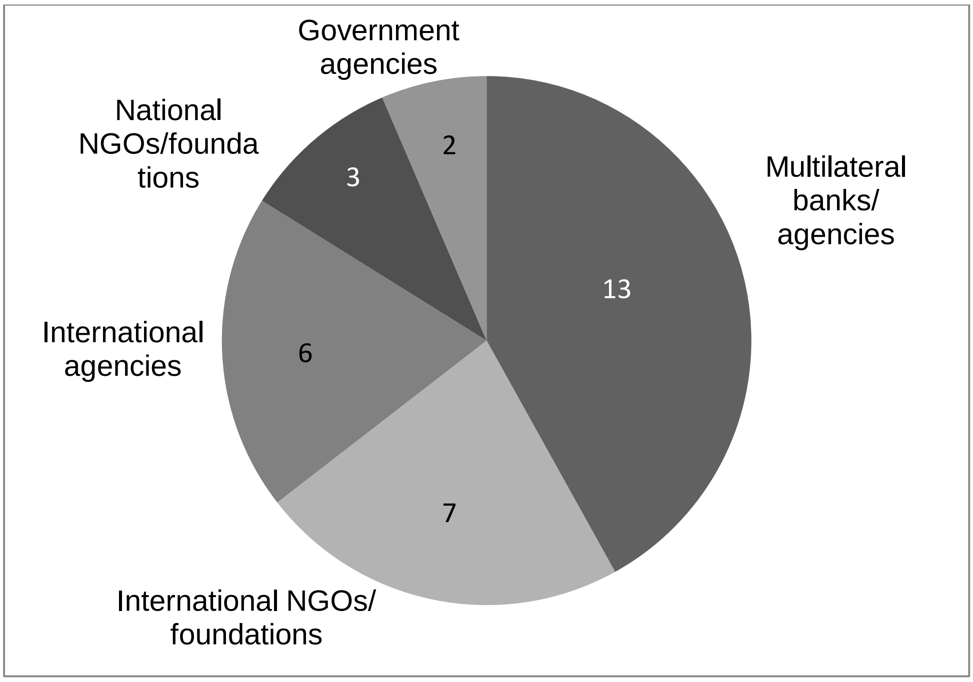
4.1 Sources of funding

As figure 3 illustrates, the vast majority of impact evaluations – 84 per cent – were financed by international and multilateral institutions: 13 by multilateral banks and agencies[5], 7 by international non-governmental organisations (NGOs) or foundations[6] and 6 by international agencies.[7] Only 5 of the studies were commissioned by national entities: 3 by national NGOs or foundations and just 2 by government agencies. Until about 2008, demand for impact evaluation came almost exclusively from international donors and multilateral banks. Since then, government

[5] Inter-American Development Bank, Development Bank of Latin America (formerly Andean Development Corporation, or CAF), World Bank and the International Finance Corporation.
[6] Academy for Educational Development, Global Development Network, Partnership for Economic Policy, United Nations World Food Programme, Think Tank Initiative and the International Center for Research on Women.
[7] International Development Research Centre, Mathematica Policy Research, the United States Agency for International Development (USAID) and the Ford Foundation.

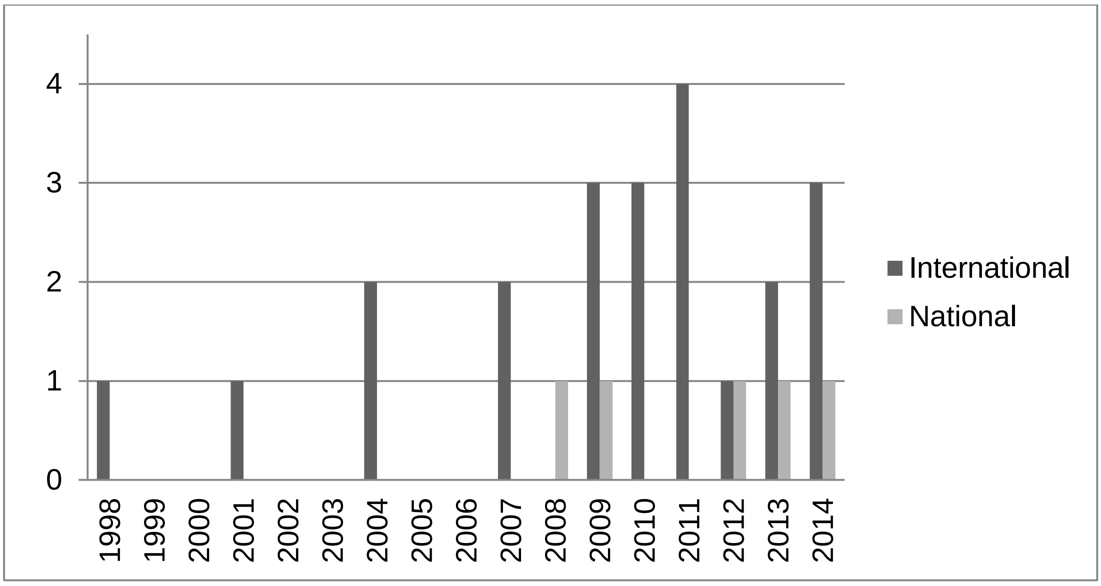
agencies and other national institutions have begun to request them, albeit still in smaller numbers (see figure 4).

Figure 3: Number of impact evaluations by type of funding agency



Note: From the GRADE projects database.

Figure 4: Institutions requesting GRADE impact evaluations



Note: From the GRADE projects database.

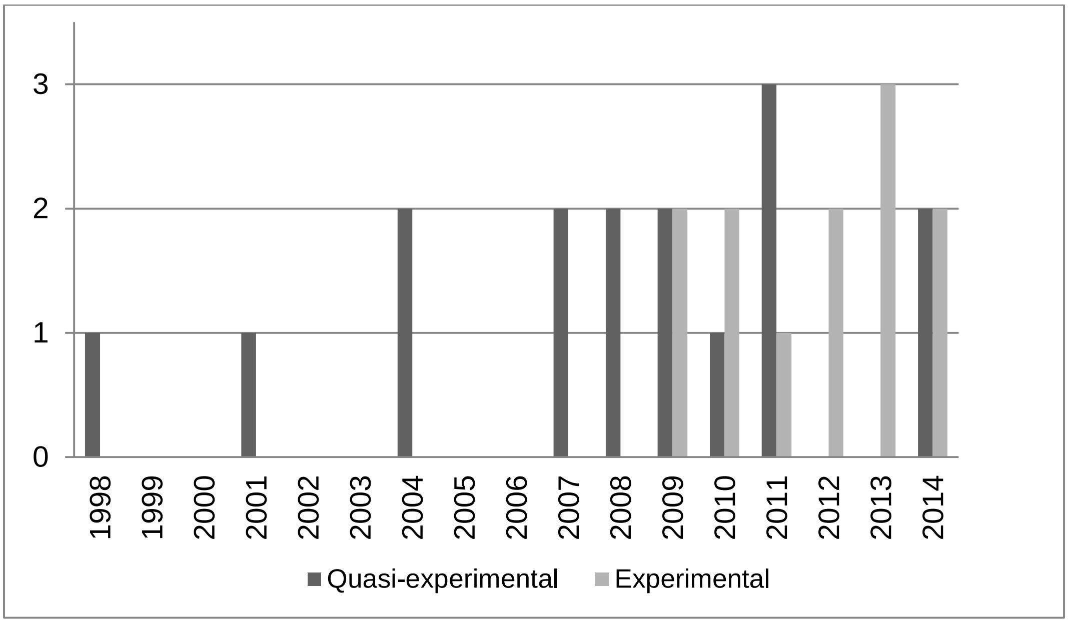
From our interviews with GRADE researchers, we know that even as interest in impact evaluation grows in the Peruvian public sector, the greatest demand is still for process evaluations. This preference is associated with the fact that process evaluations are more directly associated to the everyday tasks of public officials and programme implementers. This simplifies the preparation of research questions and the terms of reference to execute them. Furthermore, process evaluations are immediately useful in terms of providing information on the progress of programmes and policies, and they make it possible to make adjustments that can be easily implemented. Section 5 explores this point in more detail.

In the same vein, there is a good deal of heterogeneity in the demand for impact evaluations within the public sector, associated to differing capacity across sectors. At one end of the capacity spectrum, we have the social protection sector under the Ministry for Development and Social Inclusion, which has a general directorate for monitoring and evaluation, equipped with knowledgeable personnel who are capable of demanding more sophisticated and rigorous evaluations, including experimental designs. At the other end, we have sectors such as the Ministry of Agriculture, whose lack of capacity is reflected in, for example, the creation of the Mi Riego programme, with a budget of US$500 million to invest in irrigation infrastructure but no capacity to prepare the requirements of an evaluation or a baseline study that could ensure the best use of the programme's considerable resources and make decisions towards the greatest benefit for society.

4.2 Methodological designs

Until 2008, impact evaluations performed by GRADE used a quasi-experimental methodological design (figure 5). As the demand for impact evaluations grew, so did the demand for experimental methodologies (RCTs).

Figure 5: Number of impact evaluations initiated by GRADE, by methodology



Note: From the GRADE project database.

In general, the methodologies used to evaluate programmes have evolved from narrative approaches or expert opinions – from a time when these approaches were the only ones that existed – to the sophisticated methodological tools in use today. The use of baseline data occasionally accompanies these different methodologies.

Simultaneously, in the international academic field there is an important and ongoing discussion in which methodological options can be used to address the limitation (in randomised impact evaluation) of not answering a range of questions wide enough to satisfy policymakers' demands, and how to combine methodologies with other analytical tools to address the complexity of many interventions. In Peru, demand tends to concentrate on the most basic question, that is, whether a programme is having the expected effects. The more complex questions, such as why a programme fails, require a combination of quantitative and qualitative methods; narrative is useful for understanding aspects that it is not possible to address using quantitative tools.

In sum, the last two decades have seen significant changes in terms of wider use of experimental designs and increasingly localised sources of funding. The evidence indicates an increasing local demand for impact evaluation studies with rigorous designs, suggesting a bright future for the field. However, much can be done to strengthen the still-limited role of impact evaluation in informing policy decisions.

5. The influence of impact evaluation on policy in Latin America

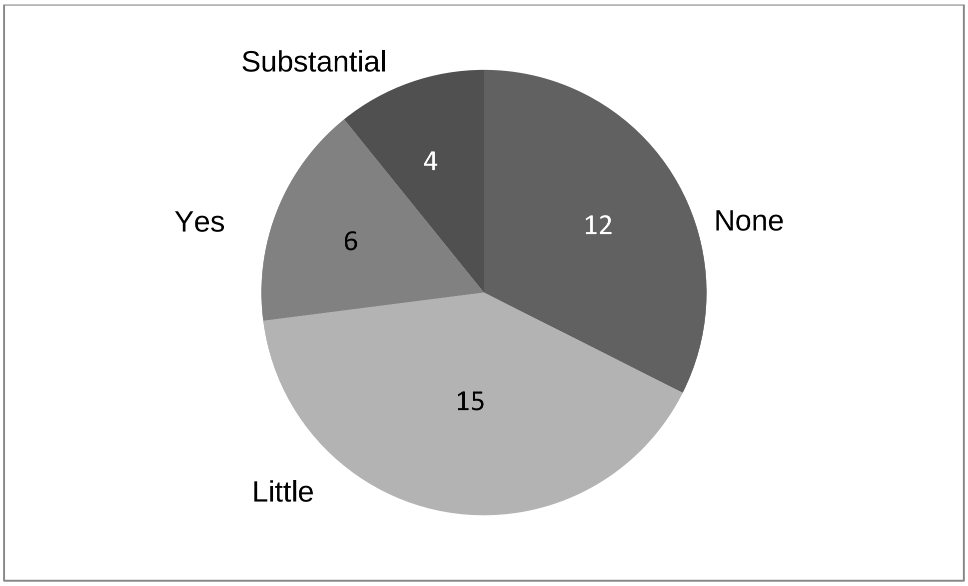
To address this question of policy influence, we use two sources of information. We first systematise and analyse the information from our survey of Latin American impact evaluation experts, as shown in the figures in this section. In section 6 we consider the information from interviews with GRADE researchers and Latin American policymakers.

We surveyed 19 impact evaluation experts based in the region, focusing on their experience with the impact evaluation process and policy influence. Altogether, they reported on 37 completed impact evaluations. Despite the small sample, the survey provides rich information that we use, not as conclusive evidence, but as a starting point to address the central question in our study.

Each researcher was asked to identify and report on the three impact evaluations he or she considered most interesting, from a policy perspective.[8] When asked how much influence their results had on programme or programme-related decisions, almost three out of four answered that their results had little or no impact (see figure 6). Only 1 in 10 reported that results had substantially influenced policy. The rest of this report explores what factors may be behind this rather modest result.

[8] Individual responses covered between one and four impact evaluations.

Figure 6: Influence of impact evaluations on policy decisions



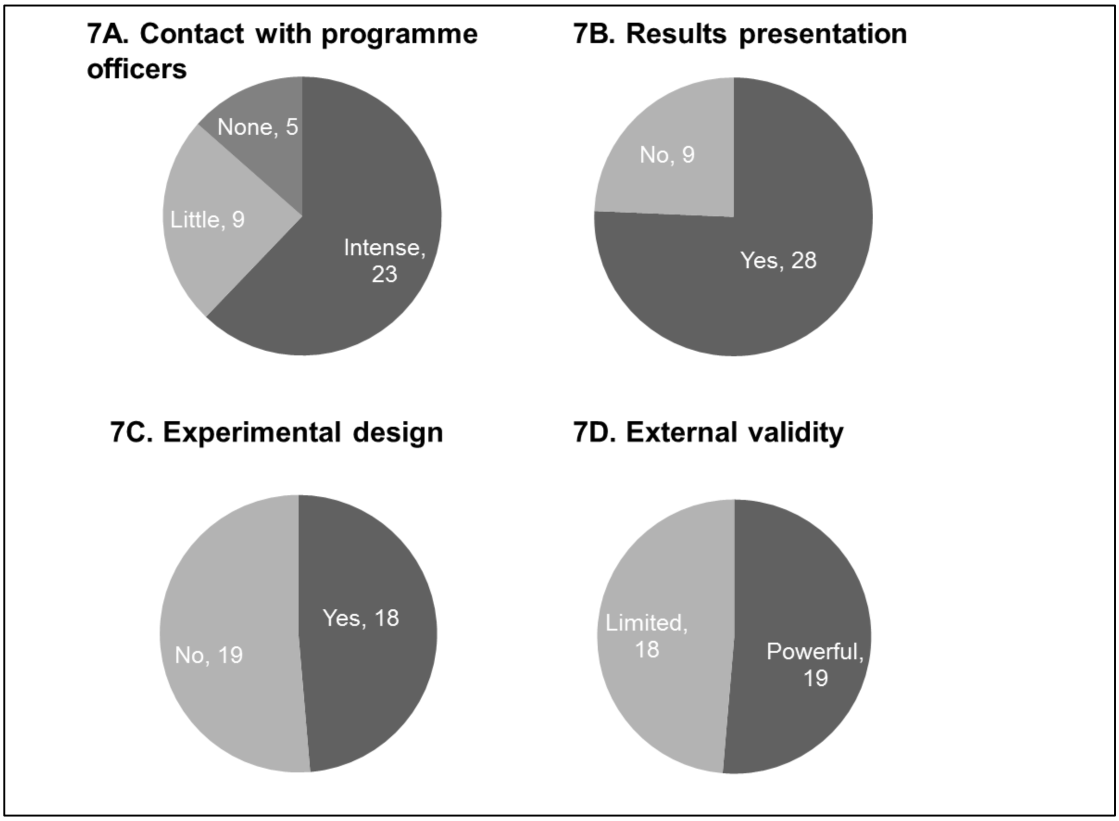
Note: From GRADE survey of Latin American impact evaluation experts.

Factors suspected of affecting the process through which impact evaluations inform policies include the researchers' degree of involvement with policymakers, the dissemination of results, the use of convincing methodologies and the external validity of the study. Figure 7 presents the distribution of responses to survey questions on each of these factors. There is significant variability across studies in all these features.

Furthermore, no single feature seems to secure attention from those making decisions. For example, intense contact between the researcher and the implementers from the design stage of the study is much more frequent than influence in decision making about the programme: three in five respondents reported having had intense contact with their counterparts (figure 7A), yet, as figure 6 illustrates, only one in four indicated that their impact evaluations had had any influence on programme decisions.

There is also much more results-presenting activity than actual influence on programme decisions. As figure 7B shows, three in four respondents reported that they had presented their impact evaluation results to the programme officers. This does not seem to reflect on the number of studies (one in four) that achieved at least some influence in programme decision making. Similar observations can be made in looking at the study's research design (Figure 7C) or the researcher's self-assessed external validity of the study (Figure 7D). We also note that most of the studies in our sample have been published; only 6 of the 37 are unpublished.

Figure 7: Factors affecting policy influence



Note: Authors' elaboration.

All of the studies that did manage to influence programme officers' decisions have at least two of the desirable features in figures 7A–D. However, having these features was not sufficient to achieve influence. As table 2 shows, only a small portion of the studies actually influenced programme decisions. Even among the nine studies that involved all three desirable features (intense contact with programme officers at the design stage, an experimental design and strong external validity, as judged by the researcher), only half achieved influence on programme decisions.

Table 2: Correlation between research features and policy influence

	Influence	No influence	Total
1. Intense contact	8	15	23
2. Results presented	8	20	28
3. Experimental design	7	11	18
4. Strong external validity	8	11	19
Two of the above features	10	18	28
(1) and (3) and (4)	5	4	9
Total	10	27	37

Beyond study-specific characteristics, the institutional context has a crucial impact on how much influence impact evaluations have on programming and policy decisions. We discuss this context and the challenges to achieving more influence in the next section.

6. Challenges for the future

Based on GRADE's experience with impact evaluation and using the information we collected through surveys (appendix B online) and interviews (appendix C online) with impact evaluation researchers from GRADE and other Latin America think tanks, we attempt to identify the challenges that impact evaluation studies face in achieving the objective of better-informed policies in the region. We also illustrate some of the main problems identified through GRADE's impact evaluation experiences in the boxes throughout this section.

A starting point is to acknowledge that research and policy design are two different processes, each with its own logic, motivation and timing. Research follows an academic logic of reflection and search for knowledge, but policy design is bound by the rationality of politics and the rigidities imposed by bureaucratic operations. For research, a primary motivation is the generation of fair and unbiased knowledge, but the motivation for policy design is the implementation of plans and procedures to generate the highest political returns, signalled by approval or popularity.

These elements present a problematic context for the shift towards evidence-based policymaking. Reforming the organisational culture within the public sector would help establish impact evaluation as part of the regular activities of any public programme, thereby granting innovation a substantial place in programme management. The underlying assumption should be that public officials and programme managers do not know everything; indeed, they cannot know everything, so they should not be punished or held responsible for less than perfect knowledge. Fostering a culture of learning and innovation seems to be an essential step for evidence-based policy to flourish, with impact evaluation demanded and used by policymakers and implementers.

Even where such culture is substantially missing, there are factors that can facilitate (or hinder) a more symbiotic relationship between impact evaluation research and the policy process. We organise these factors using the qualitative data we collected, first by characterising in more detail the disconnections between supply and demand for impact evaluations, and then by looking into the institutional factors at play.

6.1 Diverging cycles of research and policymaking

The separation between policy research and policymaking derive from their cycles. A rigorous impact evaluation requires a rather prolonged time frame to gather information and analyse data before reaching pertinent conclusions. The design of a policy cannot take nearly as long, as it needs to respond to urgent concerns, a difference that can prevent even the request for an impact evaluation.

> One month, two months. And you have the pharmaceutical companies knocking at your doors. And you have the anti-immunisation people on the other side, accusing you of wanting to kill children. Am I going to make an impact evaluation that will respond in four years? It's not going to happen. (Policymaker 1)

In addition to limiting demand for impact evaluation, the policy cycle tends to make any research design precarious. Politicians often make decisions based on funding the greatest possible number of beneficiaries, as a strategy to secure votes. This approach makes coverage the overwhelming goal for many programmes. In this context, generating evidence for more effective policies may play not even a secondary role, and research designs suffer what may amount to irreparable losses. This happens, for instance, when a group or area of intervention changes after a design was approved and a baseline constructed, rendering the baseline information obsolete and losing the comparison group (see box 2).

Another problem associated with the policy cycle is that impact evaluations are often demanded long after an intervention begins, even near the end of a project, often due to the need to fulfil a donor's requirement. In this regard, GRADE's researchers are frequently asked to conduct an impact evaluation when it is too late for a proper methodological application. In some cases the concept of 'retrospective baseline' is used to infuse some conceptual underpinning to an otherwise precarious approach to identifying the effects of the intervention.

> Under the current situation, from my point of view, they usually request impact evaluations when it is already too late. And then, since impact evaluations are requested to comply with a requirement, at the most you can expect to gather new information, if at all possible, or, as we say sometimes, to perform 'autopsies', and there is a great deal of that. (Researcher 3)

Occasionally, the intervention itself comes too late:

> In the Subsectorial Irrigation Program … which had a reasonable timeframe because JBIC [the Japan Bank for International Cooperation] demanded it, a baseline was prepared prior to commencing the project. The funny thing is that the project took some time to start; the baseline was ready some time before. The first year, the first 18 months, the project made almost no progress. The baseline was then somewhat distant and, consequently, there were serious problems to make the final evaluation. Due to all the delays in project execution, in the end we made the end line survey in 2014, five years later. (Researcher 2)

Our interviews make it clear that these divergent cycles limit the demand for impact evaluations; the processes of policy research and policymaking do not harmonise easily, or well. Even when there is will from both sides, the best-planned evaluation may run into problems because of political intervention. Under such constraints, a proper strategy for some researchers is when the impact evaluation design and methodology accommodate the policy implementation process. This implies accompanying programme implementation so that on the way, the impact evaluation design can be adjusted to capture the main features of the intervention (see box 3).

> Researchers have to work closely with policymakers in order to be able to capture the complexities of the process implementation, in its design and execution; and to be able to consider that to elaborate proper methodological strategies; not necessarily the ideal ones, but the best, given the research context. (Researcher 1)

Box 2: The problems of rigorous experimental designs and large public programmes: the impact evaluation of Pensión 65

In 2011 the Peruvian government launched Pensión 65, a national non-contributory pension plan for seniors (aged 65 and older) living in extreme poverty, without a retirement pension.

As the programme launched, a team of researchers contacted the Ministry for Development and Social Inclusion to gauge interest in an impact evaluation, which they hoped to present to a contest funded by 3ie. The objective was to analyse the effects of Pensión 65 and its interaction with other assistance programmes (formal or informal, government-run or community-based) on diverse indicators related to the well-being of senior citizens in extreme poverty and their families. The design called for a random selection of groups based on the gradual expansion of Pensión 65. After confirming the ministry's interest, the researchers presented their proposal to 3ie and received funding.

However, Pensión 65 moved with unexpected speed – a consequence of budget changes and the development of targets for nationwide coverage in a much shorter time. The evaluation design had to be changed several times to try and align it with programmatic changes without losing rigor. This created problems related to differences of implementation speed: the programme was accelerating, but the research team needed time to adjust the design and gain approval from the funding partner.

The first design change was to the unit of randomisation, from districts (municipalities) to towns with low programme coverage within the districts. The design could not be implemented, however, because in the time it took to redesign the evaluation and gain approval from the funding partner, many of the towns chosen as controls had been incorporated into the intervention group. This made it more difficult to identify an adequate control group, requiring a new design and new approval from the funder. The new design selected study and control groups based on individual assignment, with controls identified from the registry of eligible senior citizens who were not yet registered in Pensión 65.

Thus, despite interest from the ministry, the programme implementers, the research team and the funding partner, discrepancy between the rate of implementation and the rate of evaluation design hindered the study. Nonetheless, the research team has continued to work and coordinate with Pensión 65 and the funder to pursue the study.

Box 3: Collaboration between impact evaluation researchers and practitioners: the case of Haku Wiñay

> If you are arriving near the end of the intervention design process, when everything is already defined, and then you want to develop a baseline and an appropriate counterfactual condition, it is much more difficult, because you don't exactly understand the nature of the intervention and how it will be deployed. In contrast, we went in with a clear understanding. (Researcher 7)

The Haku Wiñay project ('We will grow' in Quechua) was born as part of the Juntos conditional cash transfer programme, with the objective to increase household productivity by strengthening production systems. This was to be achieved by introducing technological improvements, providing training and technical assistance, promoting inclusive business practices, and improving financial skills in Juntos beneficiary households. The impact evaluation measured the incremental effects generated by Haku Wiñay activities in areas where Juntos operated.

Interest in this impact evaluation related to the potentialities and limitations of collaborative work between researchers and implementers in a changing environment. Success hinged on the researcher's prior knowledge of the topic, his 'accompaniment' of the project and his degree of authority over the project, granted due to close ties with senior management (the researcher is a member of the Advisory Commission to the Ministry for Development and Social Inclusion, where the programme is based). These factors allowed the researcher to build trust with policymakers and practitioners and to be involved with the project from the beginning as an adviser. This also enabled him to develop a successful counterfactual condition, thanks to detailed knowledge of the assignment and deployment rules.

> The benefits from accompanying the entire process are enormous. If you come in at the right time, you know how the project has taken shape and what is going to happen and how. You know what priorities they [the officials] have in mind and everything is spelled out in detail. So everything comes together nicely. Of course, that requires that you have a position in which you came with funding and the intention to accompany the programme, which is a ton of time and work in an activity that does not offer any short-term academic gratification. (Researcher 7)

This strategy of collaborative support can be successful if it is undertaken in a disciplined manner; in this case, the researcher met with programme officers every two weeks, and these regular and frequent meetings with practitioners enabled him to influence decisions, better understand the logic of the design and ensure consistency in its deployment. It was thus possible to plan which populations to target for intervention, identify the best possible control group and randomise based on the implementation schedule for subsequent years.

6.2 Diverging motivations and uses

Impact evaluations serve three main purposes: generating new knowledge; improving public programmes, policies or interventions, including expansion or closure; and accountability, that is, informing citizens on the effectiveness and (if it includes cost analysis) efficiency in the use of public resources. Even as impact evaluation has gained space in informing public policies, and as demand for impact evaluation has increased in the last decade or so, policymakers' and researchers' expectations differ. Those creating policy are interested in showing that their programmes work; there is an overwhelming validation motive. Meanwhile, researchers are interested in the more abstract causality relationships between interventions and outcomes; indeed, the results of impact evaluations may be disruptive for public officials and practitioners. Interventions are evaluated to improve programmes – to change them. This means that impact evaluations will tend to generate innovative ideas which may clash with the world of politics, where it is hard to change pre-determined processes. A third element is the professionals whose career paths may be affected by impact evaluation results.

Thus, demand for impact evaluation is often associated with a desire to validate an intervention, the sooner the better, with little regard for technical feasibility:

> The ministry said to me: we want to have an impact evaluation performed on four programmes that have started this year. I told them not to do them because they will be shooting themselves in the foot. The purpose of an impact evaluation is to show the impact of the programme and with massive programmes that will start now and which are being prepared on the go, most likely a thousand things will go wrong. They will need to adjust them on one side and on the other, and, even if things go well, the impact will not be seen after a few months, but after a year or two. (Researcher 4)

In this context, impact evaluations that produce evidence to support interventions are helpful for practitioners and policymakers. This was the case of an innovative (at the time) study that evaluated the Peruvian government's Rural Roads programme, generating evidence on the benefits of improved road infrastructure. The study garnered interest from international entities (World Bank) and national institutions (Ministry of Transport and Communications). The World Bank received the results with enthusiasm, as they had been supporting this type of intervention all over the world. National policymakers were interested because they liked the results and wanted to know more. They did not have a clear understanding of what it took to arrive at the results, but they nonetheless used the results to ensure funding for an expansion of the programme.

Some studies reveal a lack of impact. When a programme is large enough and political stakes are high, lead organisations often try to keep results out of the public eye. This was the case of the evaluation of One Laptop per Child, which the president had sponsored. A different twist occurs when an impact evaluation shows moderate impact, but the implementing institution manipulates the results and uses them for its political marketing:

> That puts you in a very complicated situation. There was an issue of political pressure. They never wanted the results published. We noticed some impacts, but modest. The truth is that it was not a very impressive programme in my opinion … they marketed themselves as a successful programme and it became established in the popular imagination as a great programme, but the evidence of impact that we found was modest. (Researcher 4)

It is not only in the public sector that one finds this type of reaction to impact evaluation results. The case in box 4 reveals the perils of conducting an impact evaluation when there is a strong validation motive. This risk is compounded by the fact that professional advancement may be tied to evaluation results. The consequences for researchers can be dire: blocked access to the academic community means lost ability to publish and disseminate results.

From this perspective, the decision to conduct an impact evaluation should be strategic. Certainly, researchers and implementers should be wary that the impact evaluation agenda is responding to a need for better policy, not to a desire for academic recognition or a chance to contribute to a 'hot topic' in the field. One should not expect every impact evaluation to affect policies, at least not in the short to medium term.

Box 4: The perils of delivering bad news: the impact of municipal business licenses on microfirm performance in downtown Lima

In 2008, Jaramillo and Alcázar won a bid put out by the International Finance Corporation, the private sector arm of the World Bank, to conduct an impact evaluation of a programme to facilitate access to operating licenses by microfirms in downtown Lima, Peru. Since it was not feasible to evaluate the programme directly (it had ended some years before and had not collected data on its beneficiaries or comparable firms), the proposal focused on the impact of operating formally, that is, with a business license. Things went very well with the study until it produced results that suggested that the impact of licensing was nil. The researchers were not allowed to present their results, and International Finance Corporation management in Lima blocked an invitation by the International Development Research Centre (the study's co-funder) to present the results at the Canadian Economics Association meetings.

Meanwhile, the World Bank board of directors had initiated an investigation on the usefulness of the evaluations they had been commissioning. The study fell into the sample selected for the investigation and the researchers interviewed us. This brought the matter to the attention of the World Bank's board, which ordered an audit. Since programme managers had criticised the study, the World Bank hired a consultant, who looked at the databases and was able to reproduce and validate our results.

In this case, what was at stake was evidence for an investment of several hundred million dollars to streamline municipal license procedures around the world. Beyond the need to rethink intervention strategies, careers and reputations were on the line. Although the study conclusions did not suggest a need to discontinue this type of programme, a mix of frustration and lack of capacity to understand the results blocked dissemination of results and the potential for academic recognition.

Looking at positive experiences, we observe that impact evaluation researchers are able to publish documents that are used by other researchers and cited in subsequent studies. This occurs mainly with impact evaluations that generate new academic knowledge, such as the case, in box 5, of a study on microfinance and training for entrepreneurship that had a large academic impact abroad, without immediate impact in Peru. As one of the first to evaluate microfinance programmes using experimental methods, it was cited in numerous subsequent studies and used by many international organisations working on microfinance, such as FINCA and Freedom from Hunger (Karlan and Valdivia 2007). Meanwhile, it took time for Peruvian NGOs and ministry representatives to make use of the results.

Box 5: Taking the long way to policy influence: the case of impact evaluation of microfinance

Sometime around 2002, FINCA International, concerned with complementing financial services with technical assistance, contacted Dean Karlan – a researcher affiliated with Yale University, Innovations for Poverty Action and the Abdul Latif Jameel Poverty Action Lab – to conduct a study on business education for microcredit clients. Karlan was already working in this area with Martín Valdivia, a GRADE researcher. Since Valdivia was in touch with the FINCA management team, the two researchers decided to collaborate in evaluating the impact of microfinance institutions on the entrepreneurial capacity of their clients, so that the institutions could improve their businesses' performance.

One of the first documents to emerge from this joint project was 'Teaching entrepreneurship: impact of business training on microfinance clients and institutions' (Karlan and Valdivia 2006). Other important studies followed, but this 2006 paper was the most prominent, being one of the first impact assessments to use an experimental methodology (RCT) to generate new knowledge by analysing entrepreneurship among microfinance clients. The study was well received in international academic circles, and international and regional organisations (such as the World Bank, the International Center for Research on Women and the Development Bank of Latin America) indicated an interest in having the authors undertake in further studies on the topic and present and discuss their findings. Karlan and Valdivia were invited to contribute to international publications, such as the United Nations' Measuring Women's Economic Empowerment Report.

The study's success is reflected in its citation in numerous other studies on entrepreneurship and microfinance by renowned scholars such as David McKenzie, Esther Duflo and Abhijit Vinayak Banerjee. This work also generated impacts within multilateral agencies, international organisations and charitable foundations, as they adopted elements and applied them to their interventions, daily work and institutional policy. One such organisation is Freedom from Hunger, which provides business training for women entrepreneurs in microenterprises worldwide.

In Peru, however, it took a long time for local NGOs to apply findings from these studies to their work, and even longer for the findings to have a higher-level impact. The promotion of financial training based on the evidence and results from the Karlan and Valdivia studies was introduced into Peruvian public policy only in 2013 (for example, with the Haku Wiñay project). Even this step was due to the then-minister's interest and familiarity with the subject.

The lesson is that it can take years for a study of this sort to permeate public institutions and policies in a country that does not have the habit of generating evidence for policymaking. The impact of evidence may materialise only after it has had an impact on knowledge generation.

Note: A list of studies that have cited Kaplan and Valdivia (2006) is available at <http://www.mitpressjournals.org/doi/abs/10.1162/REST_a_00074?journalCode=rest#.VYSlMkZcCSp>

Positive experiences also show that despite the time it may take, impact evaluations can be critical in determining the fate of a programme, by generating evidence about whether it is producing the desired effects on the target population and ensuring that these effects are attributable to the programme. A former minister of social inclusion confirmed this role in describing two key of Peru's social programmes with opposite fates: the largest food programme, Programa Nacional de Asistencia Alimentaria (PRONAA), which shut down, and the Juntos conditional cash transfer programme, which expanded.

> They [the impact evaluations] determined their fate: for PRONAA to close it and for Juntos to boost it. They were fundamental. I would say they are the best examples we have today … When you want to close the programme, you need evaluations to tell you that you have to close it, like in the PRONAA case … But closing a programme is the extreme. If it's a programme that does not really give results, you can do more assessments to try to find out whether there is some use for it. If you do not find any, you do have to close it, that's what we expect. (Policymaker 2)

6.3 Experimental designs and the importance of complementing methodologies

Impact evaluations are demanding in terms of cost, time and technical expertise, but since the data collection is generally the main cost component, applying different methods should not create major cost differences. Experimental designs (RCTs), which are the gold standard for evaluation rigor, may involve a more careful design and thus demand greater specific expertise. On the other hand, data collection for quasi-experimental designs may be more demanding, particularly if some sort of matching is applied. In addition, although baseline data is not strictly needed in an experimental design, it increases the efficiency of the estimations and is generally quite useful to get a better grasp of the study's subjects. Between baseline and end line, the intervention must produce the sought-after effects, which typically signify a considerable wait time that is nonetheless quite valuable to policymakers.

Designing a rigorous impact evaluation is intellectually demanding, and implementation requires discipline from the researcher and the implementer. It is not always easy to find research centres with experts who are prepared to take up the cost of the study in terms of time, operating conditions and capacity, nor is it easy to find public institutions with policymakers who are prepared to do the same. In other words, a successful impact evaluation depends not only on a sound research design but also on the capacity of the entity that conducts the evaluation and the institution that manages the intervention.

In some cases, policymakers and practitioners prefer to avoid the more demanding rigorous impact evaluations and sophisticated methodological instruments. The greater cost of rigor – in terms of time, money and intellectual and operative work – leads to less sophisticated approaches as the standard for impact evaluation, but with greater acceptance.

> The main point is how the policymakers can identify ways to know when to implement an impact evaluation and what other intermediate options exist, short of an experimental impact evaluation. They must take particular care in ensuring that experimental evaluations do not generate problems in the programme's interactions with users and other actors around the intervention. Identify when an experimental evaluation may or may not be feasible. The experimental evaluation is the most powerful and strong way to measure the effectiveness of the programme, but you have some restrictions associated with the characteristics of the implementation process of the intervention itself. (Policymaker 3)

Nevertheless, it is clear that there has been progress in the acceptance and use of experimental designs in impact evaluations. In some sectors of the government, and in NGOs and cooperation agencies, the advisability of using experimental designs has been internalised. Policymakers and programme implementers in those sectors are thus more amenable to experimental impact evaluations, even when the policymaker or implementer may not know how to connect the evaluation with a programme's development goals. This public sector internalisation of the need for impact evaluation has resulted in serious, high-quality evaluations, often owing to the role played by donors and multilateral agencies such as the World Bank and the Inter-American Development Bank. However, internalisation due to external pressure has a weakness: it may come with lack of knowledge about basic concepts, such as causality or random assignment, which can create significant complications during an impact evaluation. Issues range from the ethical (Why are we denying access to the programme to some people?) to the practical (Is it possible to have an adequate control group?).

> I received a call for a program at the Ministry of Education. They were implementing an intervention to provide learning programs through video technologies to all urban secondary schools. And they wanted an impact evaluation. I asked them about the roll out of the program. They told me they were installing the technology so every school would have it by the beginning of the following school year. I asked them with who are we going to compare these schools. They did not know, but they had promised the president that the intervention would be in all urban schools by next year. (Researcher 6)

The problem is further complicated by the high rate of turnover among programme officers. At worst, this may result in a reorientation or termination of the evaluation. At best, it lengthens the study duration due to temporary paralysis and the need to train new personnel.

> Despite the level of commitment, since it is a randomised evaluation and you have to start before the beginning of the intervention, I think we are already on the eighth programme director … under these conditions [it] is very complicated.

> Interviewer: And have you had to explain again [the] randomisation?
>
> Interviewee: Every four months we had to do a presentation of the study. (Researcher 3)

Public sector programme officers prefer process evaluations, which they perceive as being closer to what they do and feel, therefore, that they understand what is being addressed, since it relates to their everyday work. The perceived distance between what implementers and policymakers do and the impact their intervention have may cause programme officers to distrust the results of an impact evaluation. On the other hand, high-level government officials are unanimous in seeing process and impact evaluations as complementary.

> For some programmes the most useful evaluation is possibly the process evaluation, because you will not close the Comprehensive Health Insurance Program with a yes-no decision. What you want to have is an evaluation that can tell you: you need to make some changes to improve the program processes or elements so as to have a better operation, a more efficient system. (Policymaker 2)

A recurring phenomenon during an impact evaluation, known as the ‘black box’ approach, is to focus on rigorously estimating the impact of an intervention and lose sight of what may be driving these results. Along that line, there is a limited supply of researchers with the capacity to perform serious, high-quality impact evaluations that avoid the black box problem. Many impact evaluations therefore limited to contrasting the change in a certain set of indicators between beneficiaries (treatment) and non-beneficiaries (control) from the beginning to the end of the intervention.

Impact evaluations will be all the more useful if their designs consider the theory of change and causality and collect information from each link or channel in the causality chain to verify why a link works in a manner other than expected (or vice versa), thus avoiding the ‘black box’ approach. In general, evaluation and learning systems require the application of different instruments to better understand interventions. Qualitative methodology is a powerful instrument for understanding causal processes and obtaining a finer interpretation of what happens during the implementation of and intervention and providing a more complete interpretation framework of the results.

> Quantitative, on the one hand is necessary, but the narrative and the qualitative data helps you understand and often, as I said, randomisation *per se* is not sufficient. (Researcher 3)
>
> In the process [of evaluation] of four or five years a lot of things happen that you may evaluate in more qualitative or process terms. You can develop evidence that allows you a better interpretation of what happens and get to your end line with a better understanding [of the intervention] and give a

> powerful interpretive framework to the effects [of the programme]. (Researcher 2)

6.4 Lack of a public impact evaluation culture and limited capacity

Institutional models to promote evaluation in Latin American public sectors do not follow a single approach. The two-pronged Mexican system separates the evaluation of social programs (overseen by CONEVAL) from the evaluation of budget programmes and policies, management processes and public services (conducted by the Achievement System Evaluation, housed in the Inland Revenue and Public Credit Secretary). In Colombia the National System of Management and Results Evaluation is run by the Department for Monitoring and Evaluation of Public Policy within the powerful National Planning Department.

In Peru, lacking another institution with sufficient power and capacity to promote and guarantee an evaluation culture, the Ministry of Economy and Finance has partly assumed this role and is doing some systematic work to promote evaluation. More importantly, the Ministry of Economy and Finance is the public institution in charge of implementing other tools that complement impact evaluation, such as results-based budgeting programmes and independent evaluations on design, process and implementation (EDEPs). It is also in charge of monitoring performance indicators and incentives programmes. Ministry officers are well acquainted with evaluation and have the budgetary capacities to set innovation in motion, but adoption in public entities is heterogeneous; greater effort is needed to promote impact evaluation and monitor whether public entities commission them and use their results to improve programmes.

> It is not a matter of 'I don't understand', 'I don't want', 'there is no budget'; the problem is that there is not an entity to enforce it and that should be MEF [Ministry of Economy and Finance], it is just that. They should mandate that every public intervention of a certain magnitude must have a baseline, it must have an intermediate evaluation, a final evaluation. (Researcher 2)

Although the ministry's promotion of EDEPs has encouraged the use of evidence in policymaking,[9] these new policies and instruments (mainly results-based budgets and EDEPs) are not enough. Proper incentives to implement impact evaluations and to further use their results are needed, but such incentives are generally missing in the Peruvian public sector. Some respondents argued for making the use of impact evaluation mandatory, as it is the case for results-based budgeting programmes.

> There should be an office within the MEF [Ministry of Economy and Finance] structure to make of this issue one of life or death. Because this implies a combination of research methods and politics, you require a certain

---

[9] Since 2008, the Ministry of Economy and Finance has commissioned an average of 8 EDEPs per year, with approximately 50 completed to date.

> technocracy that understands the method thoroughly and has the sufficient political leverage to see it through. (Researcher 2)

> The rest (sectors) only decides with bureaucratic rules and there are some bureaucratic rules that do work. If MEF's DCGP [Quality of Public Spending Department in the Ministry of Economy and Finance] limits you [with] results-based budgeting, because you have not adopted this recommendation, I assure you that [the institution] will adopt it, because there is a bureaucratic rule that gives you rewards and punishments, opens or closes doors … If you don't have the obligation to implement [then] the evaluation and the process ends there. (Policymaker 2)

Furthermore, although evaluation capacity has improved over the past decade, it is still limited and heterogeneous across the public sector. In general, because of the lack of institutionalisation in the use of impact evaluation and other evaluation instruments, there has been limited demand for personnel trained in these techniques. Some sectors have knowledgeable personnel, as in the recently created Ministry for Development and Social Inclusion. In other sectors, such as agriculture, lack of capacity is reflected in multimillion-dollar programmes that do not demand evaluations, even though evaluations could provide guidance for improving their interventions. Despite more frequent use of impact evaluations, there is limited interest, stemming not only from a generalised rejection of evaluations but also from the lack of capacities to understand them.

> They don't quite understand what these things are for. In the end, actually, my feeling is that people, in general, consider that [impact evaluations] are not helpful for policy decisions, for some reason, maybe because they don't understand them, the methodology is rather strange … [Impact evaluations] still are seen as interesting, but maybe too theoretical, or too difficult or not very useful. (Researcher 2)

There is also varied capacity within and across sectors and entities (see box 6). Even when programme officials understand impact evaluation methodologies, it does not ensure progress, because higher- or lower-level officials may have limited understanding of what was agreed to at the intermediate level of the programme.[10] This capacity gap is more common with programme officers than with higher-level officers and donor representatives.

> I would say that in general in the public sector capacities to use evidence for policy design have improved. However, they are still limited and quite heterogeneous across sectors. Ministries such as agriculture and interior are well behind in the impulse to evidence-based policy … At the other end is the Ministry for Development and Social Inclusion (MIDIS), which has a good

---

[10] For example, randomisation was understood at the intermediate level of one programme, but not at higher levels of the ministry, where there were concerns about conflict if one community is excluded and not another. In such cases, researchers need to present the study repeatedly to officials at different levels of a sector, as in the case in box 6.

> understanding of the importance and usefulness of these instruments … In a scale between 0 and 10, I would say that the Peruvian public sector is between 3 and 4 as far as use of evidence, particularly from impact evaluations, to inform policies. (Policymaker 4)

Box 6: The problem with the lack of public-sector capacity for impact evaluation: the case of the Special Project for Land Titling

In 1996 the Peruvian Ministry of Agriculture launched the Special Project for Land Titling and Rural Land Registry, which aimed to improve productivity and competitiveness in rural areas through land titling, increased legal security and reduced transaction costs for the operation of land markets and financing (with land as collateral).

The first phase, which focused on the Peruvian coast, underwent an initial impact evaluation for some of its areas using data from the third agricultural census (III Censo Nacional Agropecuario). However, there was no baseline data for comparison. The second phase, begun in 2002, aimed to complete land titling in the coastal region and start the process in the mountains. This phase called for an impact evaluation, including collection of baseline data, and a final evaluation to be completed by a GRADE research team between 2004 and 2006. The purpose was to measure the project’s effects on the titling and registration of land, focusing on the Peruvian highlands.

Implementation of the study illustrated the challenges of executing a thorough impact evaluation for the public sector and ensuring that the evaluation would be useful. The experience revealed that agriculture sector officials lacked capacity to understand the concept and purpose of an impact evaluation: they were unfamiliar with the methodology and lacked criteria for analysing and interpreting evaluation results. As a consequence, officials focused on the negative results of the intervention, paying far less attention to the contributions that the impact evaluation offered the project. There was also a lack of clarity about roles related to monitoring and follow-up of the study. Despite the GRADE research team’s efforts to train the officials – through workshops on the basic elements of impact evaluation – a lack of relevant feedback on the study and its results complicated the development of the evaluation.

Not all public officials are reluctant about impact evaluation. The concept has been incorporated in the language of the public sector, mainly among policymakers who are acquainted with the subject. Most of the time, acceptance depends on an official’s academic or professional education or experience with evaluations.

> In terms of capacities, I don’t see … many changes. I don’t know if some people know more about this, but I believe that the most important change is that the people have internalised the demands of different entities regarding the convenience [of using impact evaluation]. Let’s say, the movement has conveyed its message. In other words, a public official that accepts to have an impact evaluation conducted on the programme she manages […], how much of that is because she knows about evaluation, or how much is because she is aware of the fact that outside everyone is saying that the standard bar of good management has been raised. (Researcher 5)

There are also supply-side limitations. Peru, and the region in general, lacks a sufficiently broad academic circuit to satisfy and conceive of a market for high-quality research or to generate high-level academic spaces to produce and disseminate knowledge. This is mostly due to the fact that there are no national institutions (universities, national funds for research or private research centres) that have made a commitment to invest in this area. The lack of a high-quality academic circuit results in many national institutions and consultants offering to perform low-cost impact evaluations that do not allow for rigor and end up producing less serious results. They also seek to satisfy the ‘client’ so they can use the evaluation results in marketing, rather than to present critical results. This contaminates the impact evaluation market.

> [W]hat I don’t know is that if here, in Peru, we have sufficient academic circuits … we are just a few, there is GRADE, Universidad del Pacífico, the Catholic University … there should be that space of investment, of efforts so that all that national and also international knowledge may be systematised and properly disseminated; we don’t have that either. So I believe that is our weakest point. (Researcher 2)

One of the conclusions participants agreed to at the 2014 Second Impact Assessment Workshop on Policy and Advocacy (Sao Paulo, Brazil) was that impact evaluations are still not discussed enough, and that this is the case partly because of insufficient training. Thus, one lesson from the workshop was the following:

> The capabilities of the researcher also affect the possibility of the [impact evaluation’s] influence. Since implementing agencies have participation on various levels of the assessments is advisable to generate levels of training (for example, on methods and approaches about monitoring and evaluation). (Workshop memorandum)

In sum, although public officials have incorporated and understand the language of impact evaluation and are not all reluctant to commission sophisticated evaluations, there is still no culture of evaluation or evidence-based policy, even when there are competent officials in leadership positions. A shift towards such a culture would require norms and the engagement of qualified personnel who would assume responsibility, enforce the rules and reward public institutions that make use of evaluation results. Most of our respondents consider that a real compromise will be required to create a public sector that understands impact evaluation methods thoroughly and has sufficient capacity to engage officers and policymakers from other sectors to demand, understand and properly use impact evaluation results for policymaking. Researchers need to continue working on producing rigorous impact evaluations and promoting high-level academic spaces to produce and disseminate knowledge.

7. Conclusions and solutions for the future

Impact evaluations already form part of the map of instruments in Peru's public management, although in a much more reticent manner than researchers would hope. The GRADE experience indicates an increasing demand for impact evaluations in the last decade, and the use of experimental methodologies is gaining space, with demand coming particularly from social assistance, education and health programmes. Further evidence shows increasing demand from local sources, compared with a decade ago, when demand came substantially from international donors and multilateral agencies. We therefore conclude that the language of impact evaluation has established itself in Peru's public sector, although amidst great heterogeneity of knowledge, demand and use across and within sectors. Two factors seem to have contributed to this recent trend: public budget reform towards a results-based budgeting system and, more importantly, the presence of knowledgeable technocrats who are well versed in evidence-based policymaking.

Positive experiences include collaboration between researchers, policymakers and programme implementers to ensure pertinent and rigorous evaluation designs and fluid feedback channels to inform programme policies. Under the leadership of the Ministry of Economy and Finance, reforms in budgeting procedures that are oriented towards results rather than inputs are, if slowly, promoting a culture where evaluating and learning from evaluations is becoming less unusual. Knowledge from evaluations may take time to filter into policy, but it sometimes does.

However, a competent public sector that is able to identify and use evidence from impact evaluations to reform programmes is a prerequisite, and take-up of results to inform policymaking is still quite low. Data across a Latin American sample suggest that only one in four impact evaluation studies from the last decade managed to influence policy in any meaningful way. Even among studies that shared desirable features – such as intense contact with the policymaker or implementer from the study design stage, an experimental design and a self-reported (by the researcher) powerful external validity – only half managed to influence programmatic decisions.

A qualitative approach suggests that the evidence-based policymaking dream – of researchers providing the evidence policymakers need and demand in order to implement wiser policies – is unrealistic. Research and policymaking follow different processes, each with its own logic, motivation and timing. Consequently, research and policymaking cannot be expected to fit well or adapt to one another harmoniously. Thus, the question arises as to how to bridge the disconnections between the two cycles. Based on our qualitative data collected, we can identify several strategies:

- The first is persuasion. The idea is that researchers help policymakers understand the usefulness of impact evaluation and set up knowledge systems for continuous policy improvement: a pertinent follow-up system, a baseline that gathers relevant information or another set of instruments, such as the process evaluation. The problem is that researchers can only do so much of this while continuing to develop their academic careers, so this strategy has limited applicability. In any case, we consider this an appropriate partial solution that may be worth pursuing in some cases.

- A second strategy refers to the potential of an evaluation in which a researcher follows ('accompanies') the intervention design process and considers it in the impact evaluation design. In such an approach, the researcher must remain close to the intervention, including permanent communication with policymakers and implementers, to be able to capture the nuances of the implementation process and adjust the design and methods to arrive at the best design possible, given policy constraints. We acknowledge two potential problems with this approach. First, the final design may be far from the ideal – and may not be rigorous enough. Second, the policymaker or implementer may not be willing to tolerate so much 'snooping around' by the researcher. Like the first strategy, this one is applicable only under certain conditions in the public sector but is nonetheless worth pursuing in some cases. Ambitious impact evaluations studies of important interventions can be implemented, although these will most likely not be experimental studies.

- A third strategy acknowledges the importance of maintaining a relationship with the policymaker or implementer, while keeping a distance appropriate for the sound development of an impact evaluation. An independent intervention is performed, with government participation controlled and the state collaborating with the evaluation in a limited manner, always at a prudent distance and without decision-making capacity – mainly facilitating the target population's participation in workshops in the intervention areas – to avoid the risk of failure of the intervention. Accordingly, it is important for the sector to be interested in the initiative to ensure participation, follow-up and learning related to the development of the intervention. This is probably the most feasible and adequate strategy.

Although these strategies may fit specific circumstances, a general reaction by researchers, and one even mentioned by most of the policymakers we interviewed, is that it is not enough to try to persuade public officials about the usefulness of an impact evaluation or to find ingenious ways to get the evaluation done. It is essential to institutionalise learning systems where impact evaluations play an important role, establishing mechanisms that make their use mandatory through an entity with the authority and capacity to do it. The experience of results-based budgeting in Peru suggests that it is not enough to implement norms and rules, but that a sustained effort to establish an evaluation culture is needed. In Mexico, for instance, CONEVAL gives a prize not to the agencies that have obtained the best evaluation results but to those that have made the greatest strides in establishing an evaluation culture.

Therefore, the information and reflections we have collected through this study clearly point to a need for the public sector to move beyond the validation motive for impact evaluations and to start seeing them as learning tools. Achieving this next step will require institutional change, generating incentives to incorporate learning and implementing proper mandatory regulations to accompany the incentives.

It is also important to consider that not all public interventions require impact evaluations. The public sector may not be interested in all of the researchers' questions, some of which may not be of current interest, even if they contribute to better policymaking in the long run. Therefore, it may not be reasonable to expect the public sector to take up all impact evaluation results to inform policies or to expect all impact evaluations to have immediate bearing on public policy. Given the considerable cost of impact evaluations – in time, funds and expertise – evaluation questions should be approached more strategically. Collaboration between researchers and policymakers in formulating such questions would also benefit the cause of better policies.

Finally, some contexts are favourable for impact evaluations, such as the expansion of a programme or the design of new components for an intervention. The need to test for the potential impacts of these types of policy innovation creates the opportunity for impact evaluations to contribute to better and more evidence-guided policymaking. Finally, administrative data has been underutilised for impact evaluation, while the data programmes produced has become increasingly more detailed and sophisticated. The use of these data sets may help reduce the costs of impact evaluation, thereby furthering its use.

Appendixes

Appendixes A, B and C are available online only at the following link:

http://www.3ieimpact.org/media/filer_public/2015/09/28/wp24-appendix.pdf

References

Agosto, G, Nuñez, E, Citarroni, H, Briasco, I and Garcette, N, 2013. From impact evaluations to paradigm shift: a case study of the Buenos Aires Ciudadanía Porteña conditional cash transfer programme. *Journal of Development Effectiveness*, 5(3), pp.305–318.

Alzúa, ML, Djebbari, H and Valdivia, M, 2012. *Impact evaluation for policy making: a close look at Latin American countries with weaker research capacities*. Buenos Aires: Centro de Estudios Distributivos, Laborales y Sociales.

Behrman, J, Parker, S and Todd, P, 2009. Schooling impacts of conditional cash transfers on young children: evidence from Mexico. *Economic Development and Cultural Change,* 57(3), pp. 439–477.

Behrman, J and Sengupta, P, 2005. Progressing through Progressa: an impact assessment of a school subsidy experiment in rural Mexico. *Economic Development and Cultural Change,* 54(1), pp. 237–275.

Bernal Salazar, R, 2011. *Guía práctica para la evaluación de impacto.* Bogota: Universidad de los Andes, Ediciones Uniandes.

Bonbright, D, 2012. Use of impact evaluation results. *Impact Evaluation Note No. 4*. InterAction. [online] Available at: <http://www.interaction.org/document/guidance-note-4-use-impact-evaluation-results> [Accessed 24 September 2015].

Cameron, D, Mishra, A and Brown, A, 2015. The growth of impact evaluation for international development: how much have learned? *Journal of Development Effectiveness*. [online] Available at: <http://www.tandfonline.com/doi/full/10.1080/19439342.2015.1034156#.Ve7q_J1VhBc> [Accessed 25 September 2015].

Center for Global Development, 2006. *Will we ever learn? Improving lives through impact evaluation*. Washington DC: Report of the Evaluation Gap Working Group.

Chowdhury, A, Jenkins, A and Nandita, MM, 2014. Measuring the effects of interventions in BRAC, and how this has driven ‘development’. *Journal of Development Effectiveness*, 6(4), pp.407–424.

Díaz Langou, G, 2013. Validating one of the world’s largest conditional cash transfer programmes. A case study on how an impact evaluation of Brazil’s Bolsa Familia Programme helped silence its critics and improve policy. *Journal of Development Effectiveness*, 5(4), pp.430–446.

Gertler, P, 2011. *La evaluación de impacto en la práctica.* Washington DC: Banco Internacional de Reconstrucción y Fomento/Banco Mundial.

Jones, N, Jones, H, Steer, L and Datta, A, 2009. *Improving impact evaluation production and use*. London: Overseas Development Institute.

Karlan, D and Valdivia, M, 2007. Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions. *Working Paper 107*. Center for Global Development.

Karlan, DS and Valdivia, M, 2006. *Teaching entrepreneurship: impact of business training on microfinance clients and institutions*. Yale University Economic Growth Center Discussion Paper No. 941; Center for Global Development Working Paper No. 108.

World Bank, 2009. *Making smart policy: using impact evaluation for policy making*. Doing Impact Evaluation No. 14. Washington, DC: The World Bank.

Publications in the 3ie Working Paper Series

The following papers are available from
http://www.3ieimpact.org/3ie_working_papers

What methods may be used in impact evaluations of humanitarian assistance?, *3ie Working Paper 22*. Puri, J, Aladysheva, A, Iversen, V, Ghorpade, Y and Brück, T (2014)

Impact evaluation of development programmes: Experiences from Viet Nam, 3ie Working Paper 21. Nguyen Viet Cuong (2014)

Quality education for all children? What works in education in developing countries, 3ie Working Paper 20. Krishnaratne, S, White, H and Carpenter, E (2013)

Promoting commitment to evaluate, 3ie Working Paper 19. Székely, M (2013)

Building on what works: commitment to evaluation (c2e) indicator, 3ie Working Paper 18. Levine, CJ and Chapoy, C (2013)

From impact evaluations to paradigm shift: A case study of the Buenos Aires Ciudadanía Porteña conditional cash transfer programme, *3ie Working Paper 17*. Agosto, G, Nuñez, E, Citarroni, H, Briasco, I and Garcette, N (2013)

Validating one of the world's largest conditional cash transfer programmes: A case study on how an impact evaluation of Brazil's Bolsa Família Programme helped silence its critics and improve policy, *3ie Working Paper 16*. Langou, GD and Forteza, P (2012)

Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework, 3ie Working Paper 15. White, H and Phillips, D (2012)

Behind the scenes: managing and conducting large scale impact evaluations in Colombia, *3ie Working Paper 14*. Briceño, B, Cuesta, L and Attanasio, O (2011)

Can we obtain the required rigour without randomisation? 3ie Working Paper 13. Hughes, K and Hutchings, C (2011)

Sound expectations: from impact evaluations to policy change, 3ie Working Paper 12. Weyrauch, V and Langou, GD (2011)

A can of worms? Implications of rigorous impact evaluations for development agencies, 3ie Working Paper 11. Roetman, E (2011)

Conducting influential impact evaluations in China: the experience of the Rural Education Action Project, *3ie Working Paper 10*. Boswell, M, Rozelle, S, Zhang, L, Liu, C, Luo, R and Shi, Y (2011)

An introduction to the use of randomized control trials to evaluate development interventions, *3ie Working Paper 9*. White, H (2011)

Institutionalisation of government evaluation: balancing trade-offs, *3ie Working Paper 8.* Gaarder, M and Briceño, B (2010)

Impact Evaluation and interventions to address climate change: a scoping study, *3ie Working Paper 7*. Snilstveit, B and Prowse, M (2010)

A checklist for the reporting of randomized control trials of social and economic policy interventions in developing countries, *3ie Working Paper 6.* Bose, R (2010)
Impact evaluation in the post-disaster setting, 3ie Working Paper 5. Buttenheim, A (2009)

Designing impact evaluations: different perspectives, contributions, 3ie Working Paper 4. Chambers, R, Karlan, D, Ravallion, M and Rogers, P (2009) [Also available in Spanish, French and Chinese]

Theory-based impact evaluation, *3ie Working Paper 3*. White, H (2009) [Also available in French and Chinese.]

Better evidence for a better world, *3ie Working Paper 2.* Lipsey, MW (ed.) and Noonan, E (2009)

Some reflections on current debates in impact evaluation, *3ie Working Paper 1*. White, H (2009)

This paper addresses questions that go to the heart of impact evaluation research: how much has the field achieved towards its goal of informing public policies, and which factors may have been furthering or hindering this goal? Looking primarily at its own experience, a Peruvian think tank, the Group for the Analysis of Development, elucidates the importance of factors such as the researcher's contact with programme officers, methodological choices, external validity, institutional factors and public sector capacity, which may all play a role in making impact evaluation results more effective or less effective in informing policy decisions.

The authors find a common perception among researchers: it is not enough to try to persuade public officials about the usefulness of an impact evaluation or to find ingenious ways to get the evaluation done. The keys are to institutionalise learning systems in which impact evaluations play an important role and to establish mechanisms that make their use mandatory, through an entity with authority and suitable capacities to do so. Recent experiences in Latin America suggest that the time is ripe for this type of institutional development.

Working Paper Series

International Initiative for Impact Evaluation
202–203, Rectangle One
D-4, Saket District Centre
New Delhi – 110017
India

3ie@3ieimpact.org
Tel: +91 11 4989 4444

www.3ieimpact.org