

February 2013



About 3ie

The International Initiative for Impact Evaluation (3ie) works to improve the lives of people in the developing world by supporting the production and use of evidence on what works, when, why and for how much. 3ie is a new initiative which responds to demands for better evidence, and will enhance development effectiveness by promoting better-informed policies. 3ie finances high-quality impact evaluations and campaigns to inform better programme and policy design in developing countries.

The **3ie Working Paper series** covers both conceptual issues related to impact evaluation and findings from specific studies or systematic reviews. The views in the papers are those of the authors, and cannot be taken to represent the views of 3ie, its members, or any of its funders.

This Working Paper was written by Carlisle J. Levine and Christelle Chapoy.

About this paper

This is part of a series of exploratory papers to help develop and test a theory of change behind 3ie's commitment to evaluation project (c2e) and answer the question of how policymakers can take credit for demanding their programme to be evaluated? This is part of 3ie's effort to focus on evaluation use and find alternative incentives to generate demand led research and increase the likelihood of having a policy impact.

We hope that the discussions and consultations around the papers will help us improve our method and pressure development agencies, government departments, private aid funders and large international NGOs to improve their evaluation practices and infuse more evidence based policymaking. Updated information on the c2e project can be found at:
<http://www.3ieimpact.org/en/evaluation/c2e/>

Photograph: Via Interactive

3ie Working Paper Series Production Team: Rajesh Sharma and Kanika Jha

© 3ie, 2013

Contact

International Initiative for Impact Evaluation
c/o Global Development Network
2nd Floor, West Wing, ISID Complex,
4, Vasant Kunj Institutional Area,
New Delhi – 110070, India
Tel: +91-11-2613-9494/6885
www.3ieimpact.org

BUILDING ON WHAT WORKS: COMMITMENT TO EVALUATION (C2E) INDICATOR

Carlisle J. Levine

KonTerra Group

Email: clevine@konterrargroup.net

Christelle Chapoy

International Initiative for Impact Evaluation (3ie)

Email: chchapoy@3ieimpact.org

Introduction

The International Initiative for Impact Evaluation (3ie) was created to improve people's lives by helping policymakers ground their decisions in evidence of what works, how and under which circumstances. To this end, 3ie has funded impact evaluations and systematic reviews, co-hosted conferences, published guidance materials and disseminated evaluation findings, among other activities. These primarily supply-side interventions are helping to increase the availability of high-quality impact evaluations, as well as understanding of their utility for guiding development investments.

However, 3ie and other institutions seeking to promote evidence-based policymaking recognize that increasing demand for evaluation is as, if not more important than investing in improved supply, since the value of evaluations is in how they are used to inform decision making. It is assumed (and requires additional testing to prove) that policies based on evidence of what works will lead to better development outcomes.

3ie has recently increased its focus on evaluation use. It has been exploring alternative incentives to generate demand-led research, with the launch of a specific grant window for policymakers and NGOs. However, generating demand-led research does not completely address the challenge of use and uptake of research findings. According to Miguel Székely, when the data start coming in revealing good and bad results, "people tend to ignore evaluations as a protective mechanism."¹ So, how can policymakers take credit for demanding that their programmes be evaluated?

This question leads us to look for new ways to align the incentives of evaluation directors, researchers and policymakers. To make this leap from promoting demand-led research to promoting policymaker use of evaluation evidence, and to shift from providing individual incentives to encouraging a real cultural change, 3ie is exploring the possibility of developing a commitment to evaluation indicator.

According to 3ie, a commitment to evaluation is demonstrated when agencies commission credible evaluations on a systematic basis, making the evaluations publicly available and designing programmes that are in line with evaluation recommendations.² In this definition, 3ie is using the term "evaluation" to incorporate both monitoring and evaluation.³ This definition speaks to demand (commission), quality (credible evaluations and credible monitoring data), capacity (rigor and robustness),

A commitment to evaluation is demonstrated when agencies commission credible evaluations on a systematic basis, making the evaluations publicly available and designing programmes that are in line with evaluation recommendations.

¹ Miguel Székely, "The use of evidence in Mexico: experiences and challenges," *Journal of Development Effectiveness*, Volume 3, No. 1, March 2011.

² Howard White, Executive Director, 3ie, in a conversation with Christelle Chapoy, e-mail to Carlisle Levine, May 14, 2012.

³ On monitoring and evaluation, see Marie M. Gaarder and Bertha Briceño, "Institutionalisation of Government Evaluation: Balancing Trade-Offs," International Initiative for Impact Evaluation Working

transparency (publicly available) and use (designing programmes in line with evaluation recommendations and adjusting them in response to monitoring findings). It also places the conversation within broader discussions about good governance and demand for accountability.

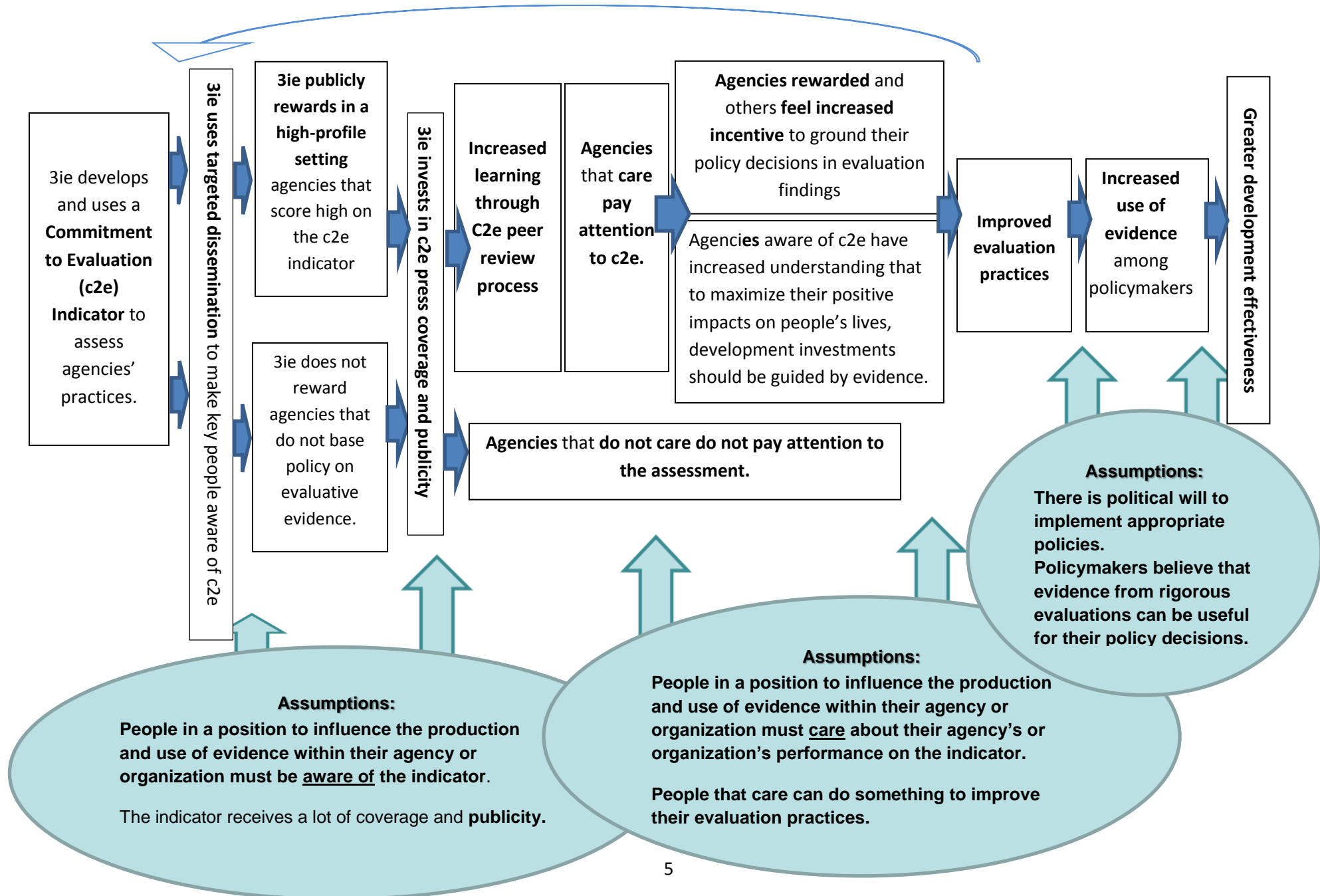
With this indicator, 3ie plans to assess the production and use of evaluative evidence by high, middle and low-income government agencies, foundations and international non-governmental organizations (NGOs). 3ie plans to recognize organizations that base policy on evaluative evidence, rewarding them for their practice, while at the same time offering an incentive to others to similarly ground their policy decision making in evaluation findings and signalling that, in order to maximize their positive impacts on people's lives, development investments should be guided by evidence. In this way, 3ie hopes to encourage the increased use of evidence and, thus, promote effective development.

For a commitment to evaluation indicator to have this desired effect, 3ie recognizes that a number of assumptions must hold. People must be aware of the indicator, care about their agency's or organization's performance on the indicator, and be in a position to influence the production and use of evidence within their agency or organization. Additionally, a link must be established between doing well on the indicator and policy changes that promote more effective development. 3ie believes it can ensure that influential people are aware of the indicator through targeted dissemination. It is the task of this paper to elucidate what might make those positioned to influence the production and use of evidence care about the indicator. The final assumption, which speaks to 3ie's *raison d'être*, requires additional research that is beyond the scope of this paper.⁴

Paper 8, July 2010, 3-4; Keith MacKay, *How to Build M&E Systems to Support Better Government* (Washington, DC: The World Bank, 2007), 7-8.

⁴ 3ie, "The Commitment to Evaluation (c2e) Indicator: A Discussion Note," January 2012, unpublished.

Graph 1. 3ie's Initial Theory of Change



3ie's creation of a commitment to evaluation indicator fits well with the current trend in development thinking. Governments are increasingly concerned with demonstrating the results of their investments. At the same time, a number of transparency initiatives, such as the International Aid Transparency Initiative (IATI)⁵ and the Open Government Partnership (OGP)⁶, have recently been launched with participation from a large number of governments, bilateral and multilateral donors, foundations and civil society organizations.

But how specifically can 3ie's commitment to evaluation indicator promote increased policymaker demand for evaluation and use of evaluation findings in decision making? Why would policymakers be interested in such an indicator and what about it would encourage them to alter their behaviour? Responding to these questions is the task of Part I of this white paper. In it, we examine the drivers behind policymakers' demand for evaluation. We then explore a variety of approaches that could potentially encourage greater evaluation demand, drawing out the theories of change that underlie them. We discuss the strengths and weaknesses of a number of models from which we can learn lessons useful for a commitment to evaluation indicator. Finally, we discuss 3ie's proposed way forward, noting factors critical to the success of this endeavour.

I Evaluation demand: What drives it?⁷

Anyone interested in what drives policymaker decision making knows that evaluation evidence tends to play a small role. Elected officials more often rely on their own experience or the positions of their political parties, funders, constituents and other interest groups to guide their decisions. Appointed officials and civil servants must consider the values, cultures, beliefs and political ideologies of their institutions, how their institutions are accustomed to operating and who has a stake in maintaining the status quo, the technical capacities of their staff, as well as other sources of information deemed credible and relevant. Both elected officials and civil servants are swayed by cost implications, as well as by the degree of change implied and therefore resources and effort – including political capital expenditures – required to adopt a new policy or implement a policy recommendation. Given the challenges of introducing change, without adequate rewards in place, decision makers will often opt for business as usual or incremental adjustments.

Evidence that informs policymaker decision making may be drawn from a variety of sources, including personal stories, national statistics and routine monitoring data. In many countries, demand for monitoring data is high at both the policy and implementation levels; it is relied upon to inform both budgeting and programme decisions. Many countries that

⁵ For more information about the International Aid Transparency Initiative, see <http://www.aidtransparency.net/>.

⁶ For more information about the Open Government Partnership, see <http://www.opengovpartnership.org/>.

⁷ For more, see Fred Carden, *Knowledge to Policy: Making the Most of Development Research* (Ottawa, Canada: IDRC, 2009), and Carol H. Weiss, Foreword, in Fred Carden, *Knowledge to Policy: Making the Most of Development Research*, x-xi; Miguel Székely, *Promoting Commitment to Evaluate* (New Delhi, India: International Initiative for Impact Evaluation, May 2012), 2-3; Keith MacKay, *Institutionalization of Monitoring and Evaluation Systems to Improve Public Sector Management*, ECD Working Paper Series - No. 15, World Bank, Washington, DC, January 2006, 5-10.

established Poverty Reduction Strategies still rely on their Annual Progress Reports to inform decision making. Many African countries issue it annually, reporting on indicators included in the national development plan. But often evaluation remains donor-driven, and evaluation data do not directly address the issues at hand in a timely, credible and easily digestible fashion. Decision makers may not be aware of the benefits of using evaluation findings, leaders may not promote their use, and institutions may not have the capacity to absorb and respond to them.

Where policymakers are more likely to demand and use evaluation evidence, the factors listed below have been found to be influential.⁸

Openness: Beyond policymaker interest and institutional capacity, studies point to the importance of “the openness of the system to the entry of new ideas [and] the democratic nature of decision making.”⁹ This might come about as a result of changes in a political system, such as the introduction of more competition between political parties or an increased balance of power between the executive branch and the legislature. Similarly, economic hardship might drive policymakers to be more interested in evidence of what works and what does not.

A champion: Where governments have made notable progress, often the role of the congress or parliament, the president or prime minister, or a central ministry such as the ministry of finance or planning is cited. Within these, having a high-level evaluation champion who can influence policy decisions and government practices can help advance the institutionalisation of evaluation.

Awareness: Some highlight the importance of policymakers understanding the utility of monitoring data and evaluation findings. Exposure to examples of cost-effective monitoring and evaluation (M&E) activities that produce actionable data or findings, and learning from the experiences of other governments that have established and are using M&E systems to inform their decision making can help increase policymakers’ demand for evaluation.

Utility: Linking an M&E system directly to government decision making processes helps policymakers perceive its utility.

Incentives: Providing incentives for using evaluation findings, such as including evaluation use in performance plans and publicly recognizing those who demonstrate good evaluation use, can help stimulate evaluation uptake.

Donor influence: When donor pressure to invest in evaluation becomes owned by a host country, then that too can contribute to increased evaluation demand.

⁸ Gaarder and Briceño, 7-8; Carden, 6; MacKay, 5-6.

⁹ Weiss, xi.

II Exploring approaches to encourage greater evaluation demand

Based on a literature review and 25 interviews conducted with representatives of governments¹⁰, the World Bank, UNDP, OECD, academia and non-profits, we identified possible frameworks for measurements and uses for a commitment to evaluation indicator. In this section, we describe them, their strengths and weaknesses and related models from which we can learn.

1. What do we mean by commitment to evaluation?

According to 3ie's definition of a commitment to evaluation,¹¹ a commitment to evaluation indicator must address five elements: demand, quality, capacity, transparency and use.

Some of these elements speak to **evaluation supply**. They ask whether credible, clear and actionable evaluation findings are available and accessible to policymakers when they need them.

- **Quality** demands an assessment of the evaluations produced and involves questions such as the following: Do the methods used produce credible data? Are the findings clearly grounded in data? Are the findings replicable?
- **Capacity** in part refers to an agency's or organization's capacity to regularly produce quality evaluations.
- **Transparency** seeks to know how easily one can find and access evaluations and perhaps also the data sets produced by them.

While examining evaluation's supply side is essential for assessing a commitment to evaluation, the measure must ultimately focus on **policymakers' use of evaluation findings** to inform decision making.

- **Demand** refers to policymakers' requests for evaluations and evaluation findings to inform their decision making.
- **Capacity** in part denotes a policymaker's capacity to understand and use evaluation findings to inform decision making and not cherry pick evidence to back up policy decisions.
- **Use** means that policymakers are using evaluation findings to inform decision making.

2. Whose commitment to evaluation are we measuring?

Clarifying whose commitment to evaluation the indicator is assessing is critical for its success. Within a government, each department or ministries whether at central or local level might demonstrate a different level of commitment to evaluation. To aggregate meaningfully the data gathered across a government's agencies would be difficult and overly burdensome for external researchers and the government assessed. This would

¹⁰ These include Colombia, Denmark, Ghana, Mexico, Netherlands, Norway, South Africa, Uganda and UK.

¹¹ Howard White.

become even more challenging to identify similar units of analysis if we include foundations and international NGOs.

For these reasons, it is important, particularly during the early stages of developing a commitment to evaluation indicator, to focus on a manageable unit of analysis, such as agencies, that still points toward systems or organizational change. By starting with a focus on agencies, rather than country governments as a whole or policymakers as individuals, 3ie can ensure that like units are being compared. This also allows the incorporation of foundations and international NGOs.

Starting with agencies and organizations that are interested in helping to develop the indicator and that volunteer to participate will help 3ie create an indicator that is indeed useful to those being assessed. If the initial group is representative of the broader range of potential targets, then the approach that this initial group designs will resonate with others when 3ie opens participation to a broader group.

A commitment to evaluation indicator must address five elements: demand, quality, capacity, transparency and use.

3. Where can we find relevant data?

The components of the indicator and how to measure commitment to evaluation will be covered in Part II of this white paper. However, as we are building the indicator's theory of change, it is important to note the challenges that will be involved in data collection. While data related to development outcomes might be publicly available (e.g., the MDGs), demonstrating a causal link between evaluation uptake and development outcomes is difficult, given the time it takes for evaluation findings to influence policy and for policy to affect outcomes. Equally difficult is separating out the contribution of evaluation uptake from the myriad factors that contribute to those outcomes. Neither task falls within the remit of this indicator. However, it will be important to articulate a clear theoretical explanation linking evidence use in policy making to development effectiveness.

Data related to the shorter-term measures of evaluation practices and use are, for the most part, not publicly available. When they are available, they are rarely comparable. The OECD DAC, the government of the Netherlands and the Centre for Global Development have all actively sought these data and found them missing.¹² Even data thought to be more publicly available and comparable, such as data related to government statistical capacity or aid transparency, have proven hard to find. As a result, the World Bank's Development Economics Data Group notes that its Statistical Capacity Index, which is based on publicly available data, is primarily useful at an aggregate level and has limited use at the country level, while Publish What You Fund has chosen to rely on soliciting responses to its own survey for its Aid Transparency Index.¹³ Given these experiences, it can be assumed that

¹² Interviews with Hans Lundgren, OECD DAC, April 2, 2012; Antoine De Kemp, Independent Office of Evaluation, Ministry of Foreign Affairs, Government of the Netherlands, April 20, 2012; and Rita Perakis, Center for Global Development, March 30, 2012.

¹³ Interviews with Barbro Hexeberg and Naoko Watanabe, Development Economics Data Group, World Bank, May 23, 2012, and Rachel Rank, Publish What You Fund, April 12, 2012.

most measures included in 3ie's commitment to evaluation indicator will require primary data collection.

4. What should be the focus of the measure?

Through our interviews, a number of possible focuses for the indicator emerged. The indicator could measure institutionalisation of evaluation or could allow low and middle-income governments to assess the utility of donor evaluation practices or both. It should also look at use of evaluation in policy making and use of evaluation for accountability. We discuss this in turn below and highlight some indicators and criteria that would be considered.

A. Measuring institutionalisation of evaluation

Measuring the institutionalization of evaluation appears to be a supply-side measure. However, since, according to World Bank research, this can rarely be achieved absent systematic evaluation use, it also points to evaluation demand.

The growing literature on experience with strengthening government M&E systems suggests that there is broad agreement among experts on a number of key lessons.... The first and foremost lesson is that **substantive demand from the government is a prerequisite to successful institutionalization**. [sic] That is, an M&E system must produce monitoring information and evaluation findings that are judged valuable by key stakeholders, that are used to improve government performance, and that respond to a sufficient demand for the M&E function to ensure its funding and sustainability for the foreseeable future.

Efforts to build an M&E system will fail unless real demand exists or can be intentionally created, especially by ensuring that powerful incentives are in place to conduct and use M&E.¹⁴

Measures of the institutionalisation of evaluation can include, among other elements, the following:

- Incentives for evaluation undertaking and use;
- The existence of a central and empowered evaluation office ;
- A law, policy, regulation or decree that mandates evaluation and its use;
- Reliable data systems;
- Structural arrangements aimed at guaranteeing M&E objectivity and quality;
- A long-term investment of human and financial resources; and
- Guidelines to conduct M&E and evaluate the M&E system itself.¹⁵

To ensure that a commitment to evaluation indicator does indeed focus on policymaker demand, 3ie may need to place greater weight on measures that examine policymaker awareness, valuing and use of system-produced monitoring information and evaluation findings. 3ie could also weight more heavily measures intended to assess how a government

¹⁴ Keith MacKay, *How to Build M&E Systems to Support Better Government* (Washington, DC: The World Bank, 2007), 53.

¹⁵ Keith MacKay, *How to Build M&E Systems to Support Better Government*, 54.

is working to stimulate policymaker demand through the establishment of an evaluation policy, capacity building opportunities for policymakers and other similar measures.

Many are concerned that it will be very difficult to design a measure such as this one that agencies will consider reliable, valid and therefore useful, since relevant and comparable data will be hard to find. Others have argued that what might be appropriate in one context might not be in another. Therefore, it would be more useful to establish measures that capture broader principles. For example, “The central location of the evaluation function close to budget and decision-making authorities” might be more appropriate across contexts than a measure that asks if an evaluation function is located in a particular office.

Nonetheless, because many government and organization evaluation directors are in the process of strengthening their M&E systems and trying to increase policymaker demand for evaluation, measuring the institutionalisation of evaluation, including its components directly related to evaluation demand and use, had broad appeal among those interviewed for this white paper. It would be particularly useful if it were connected to a peer learning component or other form of capacity building that would help evaluation directors learn how to improve areas that the measure found weak. A regional focus to a peer learning or other type of capacity building component could benefit from trust that comes with on-going relationships and would help ensure that guidance provided was relevant and applicable to the context.

Most interviewed felt that measuring the institutionalisation of evaluation, if not accompanied by a peer learning or capacity building component, would not contribute to increased evaluation demand and use. Most agencies are aware of their weaknesses. While this measure on its own might provide some additional insight, evaluation directors need assistance addressing the challenges they face in order to help agencies better institutionalise M&E systems and increase evaluation demand.

Box 1. Model – Peer learning and comparative analysis in the OECD DAC Network on Development Evaluation¹⁶

The DAC Network on Development Evaluation works to help members strengthen their evaluation systems by sharing experiences and good practice, developing norms and standards and periodically reviewing member systems. In 2009-2010, the OECD carried out their most comprehensive review study of development evaluation systems to date, based on a survey, literature review and expert interviews. Though the study drew primarily on self-reporting, findings were triangulated where possible using other literature, evaluations and knowledge of member systems gathered through the DAC peer review process (described below). Regular network meetings and established personal relationships between the OECD secretariat and member governments also helped facilitate member participation.

This study and earlier reviews are considered to be realistic assessments of OECD member evaluation systems, and the studies provide information that is not otherwise available. Members can then address the findings in discussions at network meetings, with those who

¹⁶ Interview with Hans Lundgren, OECD, April 2, 2012.

are more advanced in certain areas offering their experience to the rest. The OECD undertakes these studies periodically, not as a formal member obligation, but in response to members' expressed interest in understanding where they stand in relation to OECD norms and other members.

These studies help the OECD identify overall trends and important common issues, such as the increased involvement of partners in evaluation processes. The OECD notes that, with the pace of change in member evaluation systems, information on individual members can become out of date quickly. To provide up-to-date information between studies, the secretariat provides information on each member on its public website:

[<http://www.oecd.org/dac/evaluationofdevelopmentprogrammes/dcdndep/membersoftheda-network-on-development-evaluation.htm>].

B. Assessing the utility of donors' evaluation practices to developing country policy making

The majority of evaluation activities in many low and middle-income countries are donor-driven. Thus, increasing policymaker use of evaluation evidence in decision making depends in part on ensuring the relevance and timeliness of donor-driven evaluations. With this in mind, a number of those interviewed expressed interest in an indicator that would allow low and middle-income country government agencies to assess the utility of donors' evaluation practices to policy making related to continuing, adjusting, scaling up or stopping certain investments. At the same time, such an indicator could encourage low and middle-income country government agencies to clarify their evaluation and evidence needs. According to those interviewed, using this assessment as a basis for structuring conversations between low and middle-income country government agencies and their donors would increase the likelihood that the assessment's findings would lead to changes in practices.

In designing this measure, 3ie must clarify who within low and middle-income government agencies would participate in the assessment and on whom in donor agencies the assessment should focus. They would also need to determine the roles of central evaluation units within the assessment and follow-up conversations. If 3ie wants to aggregate the findings about donors across countries in order to make a global assessment of donor agencies and allow for comparisons among donors, then 3ie would need to ensure that the assessment examined all donors equally.

Finally, the assessment would need to be carefully worded to ensure that responses are indeed reflective of donors', rather than respondents' evaluation practices. The assessment should prioritize data-based questions and limit questions related to perceptions. Including control questions in the survey could also help overcome this challenge.

Box 2. Model – Keystone Accountability Partner Survey¹⁷

Keystone Accountability promotes constituent feedback as a means of understanding agency performance. Keystone offers four benchmark feedback surveys: Development Partnerships, Social Investors, Transnational Social Change Networks, and Grantmakers (<http://www.keystoneaccountability.org/services/surveys>).

¹⁷ See <http://www.keystoneaccountability.org/services/surveys>, accessed June 17, 2012.

All of the surveys are part of the process of improving evaluation practices through feedback. In its Development Partnerships survey, Keystone helps organisations learn how they are perceived by their partners, with the assumption that the quality of that relationship is reflective of the effectiveness of the work they are undertaking. In 2010, Keystone undertook the survey with a cohort of 28 international NGOs based in Europe and the United States who were interested in improving their relationships with their national partners, including government agencies. Questions covered financial and non-financial support, relationship and communications, monitoring and reporting. Partners responded anonymously, using an online form.

Like all Keystone's surveys, with the Development Partnerships survey, Keystone produced individual performance reports for all participating INGOs, which also let the INGOs know how they ranked within the cohort. An overall public report presented the findings about the INGO cohort as a whole. In the ideal cases, the INGOs that received their individual reports used them as a basis for conversations with their partners about steps they would take to enhance their relationships. Those INGOs then committed to repeating the survey in 24 months to see if their efforts to improve their performance had made a difference. Thirty-four INGOs have now taken the survey, which along with all Keystone surveys, is offered to organizations on an individual basis.

C. Use of evidence in policymaking

In a paper on evidence based policymaking, Miguel Székely¹⁸ reviews the different stages of evaluation use analyzed by Anderson: A first phase, where no evidence is generated in a systematic manner and policy is not based on any evidence; a second stage where scattered evaluations are used sporadically and knowledge starts being valued; and a final stage where evidence is produced and used systematically to improve social policy. Székely notes that stage two is the most commonly reached. In order to move from generating scattered evaluations to a full evaluation system that promotes evidence-based policymaking, it is essential to better understand the motives, context and profile of each actor involved in the production and use of evaluation.

Yet, interviews have revealed how little data related to evidence use are publicly available. If focusing the indicator in this way, 3ie must identify measures it can capture through primary data collection. Among others, these might include the following:

- Requirements to review existing evidence when launching a new programme;
- Existence of incentive mechanisms that promote the use of evaluation;
- Percentage of pilot programmes and programmes with impact evaluations;
- Existence of a tracking system for recommendations from evaluations; and
- Percentage of programmes that have been scaled up, discontinued, or modified based on an impact evaluation.

The last indicator in this list measures evaluation use directly. The others determine whether or not there is a system in place to promote evaluation use, making the assumption, as we did under measuring the institutionalisation of evaluation, that the existence of such a system is at least necessary, if not sufficient to bring about changes in

¹⁸ Székely M. (2011), "Toward results based social policy design and implementation", p 10.

policymaker behaviour. These measures may be assessed using evaluation officers as data sources.

Many additional direct measures of evaluation use would have to rely on policymakers as data sources. This raises the challenges of determining which policymakers to include as data sources to ensure a representative sample, as well as how to encourage their participation. 3ie would also need to ensure that it invested adequate resources in this data collection process.

D. Use of evaluation for accountability

A fourth potential focus for the indicator would be on the use of evaluation for accountability. As noted in the introduction to this paper, this focus would respond to the current global emphasis on transparency and demonstrating results. It would also complement such indices as Transparency International's Corruption Perceptions Index and Publish What You Fund's Aid Transparency Index. It would not, however, link directly to policymaker use of evaluation evidence in decision making. Rather, it would assess the availability and accessibility of quality evaluations, which is necessary, although not sufficient, for ensuring policymaker evaluation uptake.

Among some indicators that could be included in this measure are:

- The publication of evaluation reports on an agency's website;
- A public disclosure law or policy requiring the publication of all evaluation reports;
- Existence of dissemination guidelines;
- Independence of agency evaluators in terms of reporting structure; and
- Extent to which the financial and human resources of the agency are independently determined.

Most of these indicators would not be found among publicly available data. Some of these indicators could be assessed by reviewing agencies' websites. Others would require that relevant agency staff collect and share internal documents with 3ie. Others would require a survey directed at relevant agency staff.

3ie would need to dedicate the resources and time necessary for this data collection process. It would also need to devise approaches that would encourage agency staff participation. Some undertaking similar data collection processes rely on their established relationships with the agencies included in their assessments to inspire their participation. Others allow sufficient time for follow up and coaxing, sometimes using internal resources for data collection and sometimes contracting externally to assist in this process.

5. What could be the different approaches for measurement?

In terms of approaches, the indicator could be captured as a ranking, an award and/or a peer review mechanism. Each of these has different possibilities for encouraging increased policymaker demand for and use of evaluation findings. In this section, we outline the underlying theories of change behind each approach.

A. Ranking

The commitment to evaluation indicator could be captured in a ranking focused broadly on evaluation practices and use and addressing the five components highlighted in its

definition: quality, capacity, transparency, demand and use. This broad focus would recognize the range of monitoring and evaluation evidence on which agencies base their decision making.

A ranking would have multiple uses. A single number could be useful for public consumption. A multi-faceted measure could be more useful in helping those ranked understand their strengths and areas for improvement. Those ranked high could enjoy the recognition, while those ranked low might feel a greater urgency to improve their performance, although ensuring that those assessed cared about the ranking would be one of the challenges of this model. In the shorter-term, a ranking could be used to raise awareness and influence the discourse about the role of evaluation evidence in policymaking. In the longer-term, it could be aimed at influencing agency behaviour and culture change.

The effectiveness of a ranking would first depend on the credibility of the indicator. According to the experience of those interviewed, gathering data to create the ranking would be challenging. There is little publicly available data on evaluation practices and use; even less is comparable. The measure, if based on surveys or self-reporting, would be hard to validate.

The commitment to evaluation indicator could follow Publish What You Fund's lead with its Aid Transparency Index, which is further described below: as part of the measure, 3ie could ask national NGO platforms and/or evaluation associations to assess their governments' evaluation practices, and then ask governments to respond to those assessments.¹⁹ However, according to those familiar with national NGO platforms and evaluation associations, only some would be able to provide accurate information in this regard. In some cases, 3ie would need to identify other institutions capable of providing this assessment. In addition, the index may not need to be universal in terms of coverage, and participation could be voluntary initially to save cost in data collection in the start-up phase. All of these data collection challenges would open the way for those being ranked to contest the measure. Second, the ranking must garner the attention of the agencies being assessed. This could happen in four ways:

- Drawing media and/or public attention to the ranking could help increase its influence. This could be focused attention from those who care about such issues as evidence-based policymaking. Or, if the indicator were tied to discussions of accountability, transparency and good governance, then it might draw more mainstream media and public attention.
- If the ranking could become a reference in donor decision making or other investment decisions, then its influence could be even greater. In this case, 3ie would need to demonstrate the linkage between evaluation use and better development outcomes – a linkage that could be hard to establish, as noted earlier.
- Tying this ranking to an award with high visibility (see below under Award) could also increase agencies' awareness of it. This approach would be the most positively focused of the four options for raising awareness of the ranking.

¹⁹ Interview with Rachel Rank, Publish What You Fund, April 12, 2012.

- Using rating bands, which would allow agencies to be classified within four to five different levels, such as nascent, developing and mature, rather than an agency by agency ranking. This group based approach would be less of a disincentive to low ranked institutions which may face capacity constraints in getting started.

However, simply raising the ranking's visibility is not sufficient to guarantee that it will encourage agencies to invest in evaluation demand and use. Those ranked high and publicly rewarded might be happy, although caring about the ranking would depend on the importance the agencies give to 3ie and the credibility they attach to the measures that make up the ranking. Those ranked low might ignore the indicator or criticise its composition. If moving up in the ranking seemed impossible to achieve, they might be discouraged from investing in evaluation and evaluation use.

The approach that interviewees thought would most likely link a ranking to behaviour change would be to tie a ranking deemed credible by those assessed to a peer learning effort. A multi-faceted assessment could highlight for agencies their strengths and weaknesses. A peer learning effort could help them address their weaknesses by learning from others' experiences. A peer learning component is addressed in greater detail below.

Table 1: Summaries of different examples of Indices highlighting the main objective and key elements for each model. A detailed narrative is available in Annex 2

Models	Objective and key elements
Corruption perception index <i>Transparency International (TI)</i>	Objective: Developed as an awareness raising tool Key elements: <ul style="list-style-type: none"> • Made up of publicly available data from 17 reliable and recognised sources • National TI chapters use the index in their advocacy, raising public awareness • TI correlate the index with the human development index and other measures to call attention to how corruption affects development outcomes • CPI is not comparable year to year on a country basis as the methodology changes and scores are recalibrated • TI developed a number of complementary tools and more in-depth research on national level corruption. • Receives a lot of media attention
Quality of Official Development Assistance (QuODA) Index <i>Centre for Global Development and Brookings Institute</i>	Objective: Complement the aid monitoring through the Aid Effectiveness process with an independent qualitative assessment. Key elements: <ul style="list-style-type: none"> • Use publicly available data that assess donors across four dimensions • The four dimensions are intended to give donors specific enough information about their strengths and weaknesses so that they will know what actions they need to take to improve their practices • Developed by two well established organisations, respected for their scholarship and well regarded for their independence

	<ul style="list-style-type: none"> • Received media attention • Made presentations to some of the ranked agencies at their request • Some donors are using it as part of their reviews of their own practices • Use OECD data, which are a few years old and limits the short term utility of the index
Doing Business Index <i>World Bank and International Finance Corporation</i>	<p>Objective: to provide an objective basis for understanding and improving the regulatory environment for business around the world.</p> <p>Key elements:</p> <ul style="list-style-type: none"> • Gather and analysis quantitative data related to business regulations and their enforcement for domestic small and medium sized companies. It includes 11 business regulation topic areas 9 each represented by an indicator set) in 183 countries • Prior to the launch only perception based data were available • Partnered with academics who wrote background papers for each indicator. These papers were published in peer reviewed journals. • Requires primary data collection. Around 40 team members work with 9,000 professionals and government officials • Received media attention • Benefit from the World Bank credibility and reputation • The team prioritised donors buy-in and on-going discussions • Data disaggregated down to processes within a regulation topic, so that governments can understand their strength and weaknesses • Collected some data about usage: the index stimulated policy debates in at least 80 countries; around 30 governments have formed regulatory reform committees that report directly to the President or inter-ministerial level; governments have reported more than 300 regulatory reforms implemented with inputs from the Doing Business Index
Aid Transparency Index <i>Publish what you fund</i>	<p>Objective: to capture and influence the practices of their advocacy targets</p> <p>Key elements:</p> <ul style="list-style-type: none"> • Developed their own data collection instrument, comprised of a survey and two other data sources. The survey is sent to national NGO platforms to assess donors and invite the donors to feedback on their ranking and provide inputs • Recognise donors who are performing well with positive press releases • Conduct regular power mapping to identify donors that can influence change and focus their communications accordingly • Actively encourage donors to engage with them and provide inputs and feedback • Conduct indirect advocacy by briefing other civil society organisations on certain issues • Pays a lot of importance on how the index is publicised and has developed an interactive website and specific engagement strategies for each donor

<p>SmartAid for Microfinance index <i>Consultative Group to Assist the Poor (CGAP)</i></p>	<p>Objective: help agencies assess and benchmark performance, provide incentives to make improvements, and further stimulate the international debate on the quality of aid</p> <p>Key elements:</p> <ul style="list-style-type: none"> • Index emerged from the peer reviews led by CGAP in 2002 Peer reviews built relationships and ensured that organisational systems were thoroughly assessed (but time consuming). The reviews identified success factors and constraints to applying good practice in microfinance and provided concrete recommendations in Letters to Management for each agency. Through these actionable Microfinance Donor Peer Reviews, CGAP distilled in a simple yet powerful analytical framework of five key elements of effectiveness for donors to support microfinance, which formed the basis of the indicators developed in the second phase. • The index was developed out of the peer review experience, the review of the measures and a series of consultations. • The Index depends on examining internal documents shared by participating organizations. Over time, the SmartAid team has learned what documents organizations have, which are most helpful, and how to ask for them. • Participating organisations upload documents directly onto the smartAid platform • This data collection exercise is turned into participating organisations' own learning process and is more sustainable • Risks missing aspects of participants' systems that are not capture in documents • Participating organisations receive a report detailing strength and weaknesses, highlighting good practices, offering recommendations and providing an overall score • Provides an opportunity for peer learning • Serves as an external validation for the changes that members have already identified and highlights changes members may have overlooked • CGAP has a biennial assessment period where CGAP staff members allocate up to half their time to assess around 20 organisations in one round • CGAP does not seek media attention • CGAP does not make it mandatory for its members to publish their reports. Participations is voluntary
--	---

B. Award

The commitment to evaluation indicator could form the basis for a regularly presented award. This award would recognize the achievements that those assessed were making in their evaluation practice and use. The intent would be to encourage those recognized to continue their investments, while serving as models for others.

Some of those interviewed said their governments would appreciate receiving positive international recognition within a broader forum. However, the value placed on the award would depend on how the practices rewarded are identified and defined, how well the award is designed, who gives the award and how important that person is, the importance of the

ceremony, and the amount of international recognition it could confer. The award would need to be linked to a high-level international policy event and involve high-profile policymakers and renowned experts and economists. A monetary award may or may not have a positive effect. It could be useful to traditionally underfunded entities. However, its sustainability would be questionable.

3ie would have to carefully define what sorts of entities could be eligible to receive the award: individuals, agencies or governments. An award could unintentionally incentivize an individual without incentivizing an institution or system. Identifying whom the award targets will determine who will get excited about it and be influenced by it.

Box 3. Model – CONEVAL Award for Good Practices in Monitoring and Evaluation²⁰

The Mexican Government's National Council on the Evaluation of Social Development Policy (CONEVAL) was created by the Mexican Congress in 2005 to generate objective information about social policy and poverty that can be used to improve related policy decisions.²¹ The same law that created CONEVAL made evaluation mandatory for social policy programmes in Mexico.

In 2010, CONEVAL instituted an Award for Good Practices in Monitoring and Evaluation to recognize the advances institutions and programmes are making in this regard. The award recognizes achievement in a number of categories, including impact evaluation, evaluation use and logframe development, to reflect the breadth of monitoring and evaluation investments that institutions and programmes are making. The award consists of public recognition with the press and sometimes also supervisors present; the institution or programme recognized receives a trophy, while team members receive certificates. For the second year of the award, CONEVAL videotaped award recipients describing their work and noted the pride with which they did so.

C. Peer review

Those interviewed about the design and utility of a commitment to evaluation indicator expressed the most interest in an indicator that would be linked to peer learning or other capacity building opportunities. For them, being informed about their strengths and weaknesses in encouraging evaluation demand is not sufficient. They then want to be presented with opportunities to address the weaknesses that the indicator has identified.

In order to ensure that the approach helps to increase policymaker demand for evaluation, it might not be wisest to target the policymakers themselves. Participation in peer learning experiences is limited, and selecting individual policymakers to participate could result in strengthening their individual demand for evaluation without affecting the broader institutions and systems in which they work. In addition, by targeting policymakers directly, the peer learning events might draw in those who already recognize the value of evaluation and miss others entirely. By targeting heads of evaluation units who are charged with increasing evaluation demand in their organizations and/or governments, on the other hand, 3ie could contribute to institutional and system change. Peer learning would give

²⁰ Interview with Gonzalo Hernandez, CONEVAL, April 9, 2012.

²¹ See <http://www.coneval.gob.mx/cmsconeval/rw/pages/quienessomos/creacion.es.do>

these evaluation heads the opportunity to learn from the experiences of others who are trying to achieve the same in different national contexts.

Given that much of the evaluation that takes place in low and middle-income countries is donor-driven, this approach could help donors understand how evaluation might be made most useful to governments receiving their investments. At the same time, it could provide low and middle-income countries with additional models for how evaluation demand can be encouraged.

In the examples studied, peer learning has been found to be most effective when participants are able to establish trust within the group and have reason to maintain on-going relationships. In fact, in some examples, the informal exchanges that participants initiate themselves after a peer learning event can be as, if not even more beneficial than the peer learning event itself. Peer learning events can help build the relationships that make these informal exchanges possible. As such groups that participate together in a peer learning event should have some commonalities, whether that is geographic or topical interest.

Managing peer learning events, especially as numbers of participants increase, is very demanding. 3ie might consider focusing the events regionally to reach all who might be interested, while keeping the numbers of participants reasonable. 3ie also might consider partnering with already existing platforms and/or another organization in this effort.

Box 4. Model – OECD Development Assistance Committee (DAC) peer review process²²

The OECD DAC peer review process is intended to encourage dialogue and learning among member states. It involves a review of the entire development cooperation system, including evaluation, for each DAC member on a rolling basis. Peer reviews are meant to spur collaboration and change, and are also used to take stock of current policy and practice. Its effectiveness has been determined to depend on a combination of factors: value sharing and mutual trust among participants, member commitment to and sense of shared ownership for the process, as well as the credibility of the peer review process itself.²³ Participation in peer reviews is a requirement of all DAC members. The assessment of member evaluation systems is based on internationally agreed norms and standards, including the Principles for Evaluation of Development Assistance (OECD 1992) and the Quality Standards for Development Evaluation (OECD 2010). The peer review tool is available on the OECD website:

<http://www.oecd.org/development/evaluationofdevelopmentprogrammes/dcdndep/41612905.pdf> .

²² *Peer Review: An OECD Tool for Co-operation and Change* (Paris, France: OECD, 2003), <http://browse.oecdbookshop.org/oecd/pdfs/free/0103083e.pdf>, accessed May 28, 2012; interview with Hans Lundgren, OECD, April 2, 2012.

²³ *Peer Review*, 19, and interview with Hans Lundgren.

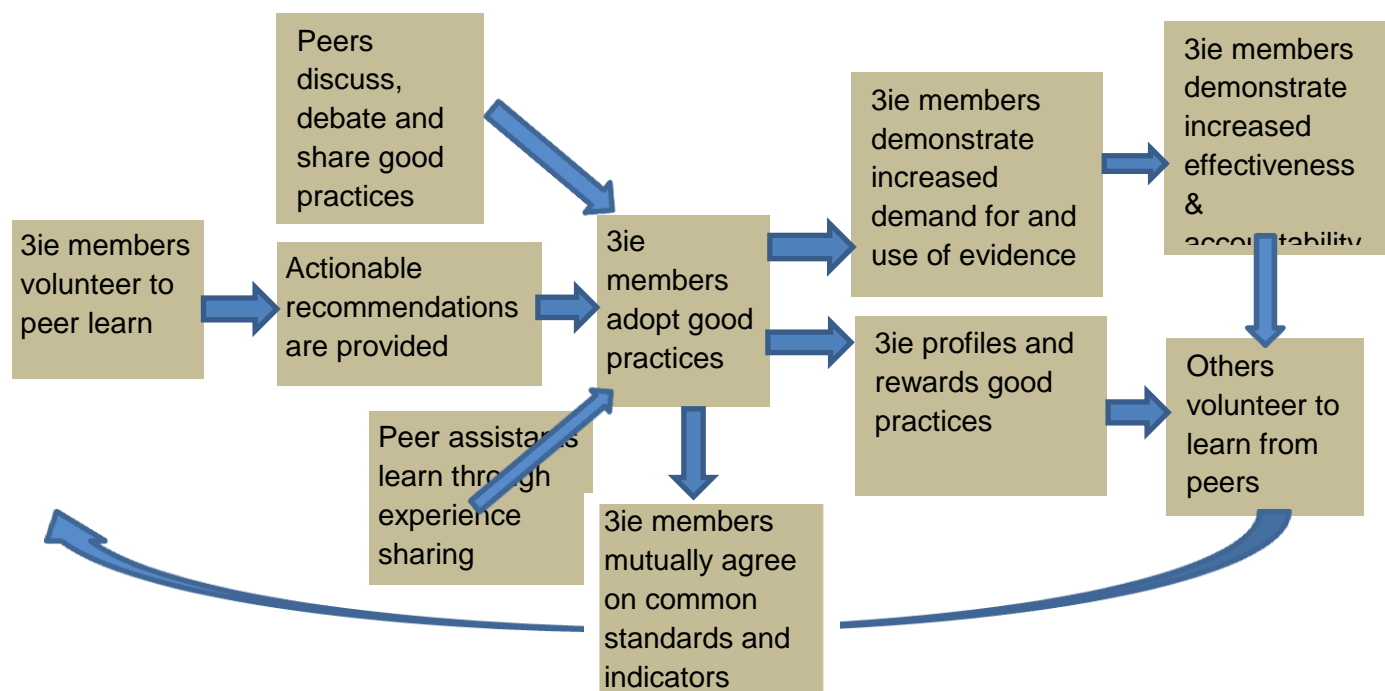
III 3ie's proposed way forward

As noted at the beginning of this paper, for a commitment to evaluation indicator to contribute to increased evaluation demand and use, people in a position to influence the production and use of evidence within their agency or organization must be aware of the indicator and care about their agency's or organization's performance on it. 3ie will use targeted dissemination to ensure that influential people are aware of the indicator. In this paper, we have explored what it will take to make these influential people care about the indicator. This analysis has guided 3ie's thinking on a way forward.

Based on the interviews undertaken and the literature review, a commitment to evaluation indicator will most effectively contribute to increased evaluation demand and use if it is tied to a peer learning process in which those charged with raising evaluation demand and use in their agencies can learn from others' experiences. As such, in the multi-pronged approach to the indicator that 3ie is proposing, the first phase will be the establishment of a peer learning mechanism with 3ie members – a group of low, middle and high-income country governments, multilateral donors, foundations and international NGOs who are already engaged and committed to the rigorous evaluation of the programmes they support – which will inform the design and development of the commitment to evaluation indicator in a second phase. By focusing on its own members, 3ie will overcome the challenge of establishing its legitimacy to create the indicator. By choosing to focus on institutions, rather than national systems, 3ie has clarified whom it is targeting, defined its targets so that it will be comparing similar entities, and eliminated the problem of aggregating large amounts of data within each country. 3ie may expand the use of the indicator to examine broader national systems once the pilot phase is complete and the indicator has been fully tested.

The objectives of the pilot peer learning mechanism are to: first, promote learning of good evaluation practices and methods; second, create a non-threatening, reciprocal and accountable system that highlights gaps in evaluation practices; and finally, generate a set of common principles and standards on the basis of which peers can assess the production and use of high quality evidence. Through this process, 3ie and its members will agree on and test the components of a commitment to evaluation indicator.

Graph 2. Theory of change of the peer learning mechanism



The interviews and literature review also showed that evaluations that are currently taking place in low and middle-income countries tend to be donor-driven and vary in their utility to those country governments. Where governments found donor-driven evaluations helpful and were more likely to make use of the findings, the governments were involved in the evaluation designs and implementation. In some cases, the governments subsequently adopted the donor evaluation approaches as their own and have used them to continue generating evaluation findings that inform their decision making.

In order to help more low and middle-income governments benefit from what are currently often donor-driven evaluation practices, another component of the pilot peer learning mechanism will involve low and middle-income country governments represented among 3ie's members to participate in the peer learning of the donor members. This assessment will help frame conversations between low and middle-income country governments and donors regarding how donor-driven evaluation practices might most effectively support low and middle-income governments' development processes. By designing and using a mutually agreed upon approach to undertake this assessment, the pilot peer learning mechanism will also inform the design of the commitment to evaluation indicator in a second phase.

UNDP has expressed interest in this model and could explore the possibility of testing some of the peer learning criteria and standards for its Assessment Development Results initiative. A draft set of criteria and standards are presented in Annex 1.

If this pilot phase is successful, then that very success might attract others to join the effort. Alternatively, participants could invite their peers to join. At that point, 3ie will have

a created a number of tools necessary for the effectiveness of its commitment to evaluation indicator:

- A tested and validated commitment to evaluation indicator;
- An instrument capable of assessing an agency's institutionalisation of evaluation, including its evaluation demand and use;
- Established processes for peer learning about institutionalising evaluation and increasing evaluation demand and use;
- An instrument that allows low and middle-income country governments to assess the utility of donor evaluation practices for low and middle-income country governments' policy making processes; and
- Established processes for undertaking structured conversations between donors and low and middle-income country governments about the utility of donor evaluation practices for low and middle-income country governments' policy making processes.

3ie's challenge will be maintaining the quality and utility of the endeavour while expanding it to include larger numbers of institutions. 3ie might consider partnering with other organizations and/or existing regional platforms to help it take this effort forward.

As a final component to the indicator, 3ie is considering creating a global award for good evaluation practices. 3ie would need to target the award carefully in order to effectively promote evaluation demand and use in a way that brings about institutional or systems change. Targeting policymaking teams, offices or agencies, rather than individuals, would be important. To identify candidates for the awards, 3ie and its partner organizations could announce a call for nominations using carefully defined criteria. Marketing the award will be essential to ensure that all potential winners are aware of the award and have a chance to submit their nominations.

The award would have to be carefully designed to give it credibility. Offering multiple awards could recognize the variety of components important to a system that is committed to evaluation, while at the same time making the award seem attainable to those whose interest in it 3ie is trying to generate.

3ie would contribute its technical expertise, eminent researchers and high quality review process, which would grant stature to the award. Offering the award in partnership with another high-profile organization could further increase its profile and legitimacy. In addition, 3ie would need to identify a venue that would give the award importance and offer recipients international recognition. Exploring media links could further help draw attention to the recipients' good practices.

Conclusion

Those interviewed regarding this project offered both warning and direction for how 3ie might use a commitment to evaluation indicator to increase evaluation demand and use. Having reflected on that input, 3ie is beginning to design a multi-faceted approach for such an indicator that has good potential for achieving its aim. Most important as 3ie advances this effort is to continue the broad consultations that 3ie is already undertaking. The input received will help ensure the effectiveness of the indicator design, and the broad outreach will help increase buy-in to the effort.

Annex 1. Suggested criteria for the peer learning

Challenge in data collection: Interviews have highlighted that evaluation data were largely not publically available, and where they are available they were not comparable. Several raised the fact that use of evaluation results and quality of evaluation could not be captured quantitatively. This measure will require primary data collection.

Methodology and data sources: Similar to the CGAP's SmartAid for microfinance index, the indicators will emerge from the peer learning mechanism. The reviews will help distil the key components for an analytical framework, which will form the basis of the indicator to be developed in a second phase.

The assessment for the peer learning will be primarily based on: (i) a questionnaire completed by NGO platforms (such as BOND and Development Initiative in the UK; InterAction and CGD in the US) or counterparts from peer departments; and (ii) internal documents submitted to the Secretariat including annual report, evaluation policy, framework and guidelines, evaluation work plan and/or strategy, CVs of evaluation staff and others. The results of the survey will then be shared with the *learning peer* who will be invited to comment and provide additional inputs. The peer assisting team will then assess the evaluation practice of the member during a field mission and consolidate the analysis in a report to be shared with the member and discussed during a peer learning session at the Members' annual conference. The *learning peer* will be consulted throughout the peer learning exercise.

Key references

Paris Declaration on Aid Effectiveness (2005) and Accra Agenda for Action (2008)
Principles for the Evaluation of Development Assistance, DAC (1991) reconfirmed in 1998 and Evaluation
Systems and Use – a Working Tool for Peer Reviews and Assessments (2006)
Managing for Results: A Guide to Using Evaluation in the United Nations Secretariat (2005)

Criteria: A first list of criteria is presented below.

Policy and practice

- Existence of an evaluation policy or an overall evaluation plan for the organisation (i.e. What is the evaluation policy? How is it made known within the organisation and how is an evaluation culture promoted?);
- Requirements to conduct rigorous evaluation (i.e. does the organisation require that impact evaluations are conducted for pilot programmes? Or have specific rules when a process evaluation is conducted versus an impact evaluation?);
- Number of impact evaluations commissioned/planned this year;

Evaluation use

- Requirements to review existing evidence when launching a new programme (i.e. when launching a new programme, are the programme managers required to look at existing evidence? Is the project document referring to existing evidence in the business case?);

- Existence of incentive mechanisms that promote the use of evidence (i.e. career incentives, reward systems, budget scoring or separate allocation);
- Percentage of approved programmes with an impact evaluation;
- Existence of a tracking system for recommendations from evaluations (i.e. How is the management response to evaluations and its follow-up organised?);
- Percentage of programmes that have been scaled up, discontinued, or modified based on an impact evaluation;

Institutionalisation

- Existence of an independent evaluation office or commission (i.e. Size of the evaluation department and scope of its mandate);
- Technical capacity of evaluation staff;
- Existence of a legal mandate for evaluation (i.e. existence of a broad legal mandate for evaluation; existence of norms including guidelines and models for standardized TOR for evaluations);
- Strength of database systems (i.e strength of the data collection and processing system);
- Strength and quality of the M&E system (i.e does the organisation have a relevant and sustainable M&E system?);
- Extent to which impact evaluations are undertaken in the agency;

Transparency

- Requirement to publish all evaluation reports (i.e. guarantee of full public disclosure through legislation on access to public information or transparency);
- Existence of dissemination guidelines;

Independence

- Reporting structure for evaluations commissioned (i.e. governance and location of the department or office; who does the office/commission report to? Who identifies priorities for the evaluation plan? To what extent is the evaluation process independent of management and political influence?);
- Extent to which the financial and human resources of the agency for evaluation are independently determined;

Financial capacity

- Budget allocation for generating and using evidence (i.e. estimated cost of all evaluations conducted in a year; our financial and staff resources dedicated to evaluation adequate to meet objectives?);
- Budget for evaluation department;

Human capacities and skills

- Ability within staff to undertake independent evaluations that have the potential to influence policy and programs (i.e How are the DAC quality standards for evaluation used?);
- Extent to which evaluations are outsourced to outside consultants/agencies.

Annex 2. Detailed narrative for examples of indices

Model – Transparency International’s Corruption Perception Index (CPI)²⁴

Transparency International (TI) first published its Corruption Perception Index (CPI) in 1995 as an awareness-raising tool. The CPI’s primary aim was to bring corruption, a topic which was at the time little discussed, into the international conversation. Because this simple tool, comprised of a number and rank and complemented by global messaging and creative infographics, is easy to understand and absorb, it became a useful entry point for journalists into the topic of corruption at both international and national levels. While the CPI is built using a solid methodology (it is made up of data from 17 reliable and globally recognized sources), it was the broad media attention that the CPI received that made it effective. The media attention led to public, donor and business attention, with the first using it in their advocacy vis-à-vis their governments and the latter two using it to help inform their investment decisions.

As a result, governments that are assessed by the CPI pay attention to it. Originally, the CPI ranked 54 countries; as of 2011, it ranks 183, treating high, middle and low-income governments equally. Those that are already committed to reducing corruption, such as those at the top and high performers in various regions, take pride in their rankings. For them, the CPI serves as an additional incentive. Those less dedicated to reducing corruption complain about the CPI and its methodology, design their own measures, or simply try to ignore it. Yet, in those cases, the CPI still serves as a useful tool. National TI chapters use it in their advocacy, raising public awareness that puts additional pressure on these governments. Internationally, researchers are able to correlate it with the Human Development Index and other well-established measures to call attention to how corruption affects development outcomes, further influencing donor investments and pressuring lower-ranked governments. Thus, the CPI influences the governments it ranks both directly and indirectly.

TI has continued to improve the CPI methodology over time. Since scores have been based on country rank in the data sources, the CPI is not comparable year to year on a country basis, based on a country’s score. Similarly, some countries enter and leave the ranking, as a result of data availability in a given year, causing other countries to move up and down in the ranking based on their exit and re-entry. Thus, one cannot compare a country’s absolute ranking from year to year. What is useful to compare is a country’s ranking relative to other countries, since this has not been affected by TI’s adjustments to the CPI methodology.

TI has achieved its primary aim with the CPI: to get corruption into the international conversation. Since the creation of the CPI, TI has developed a number of complementary tools, such as the National Integrity Systems Assessment, and has begun undertaking more in-depth research on national-level corruption.²⁵ These efforts are producing additional

²⁴ Interviews with Peter Eigen and Deborah Hardoon, Transparency International, April 4 and June 20, 2012.

²⁵ See <http://www.transparency.org/research>, accessed June 20, 2012.

analysis that TI and others can use to advocate for increased government integrity. With this, TI has questioned the ongoing utility of the CPI. In its assessment, TI has determined that the CPI is worth continuing because of its contribution to the international conversation, appropriateness for the media, and utility for researchers interested in correlating corruption data with other development-related data. Further, as part of TI's commitment to the continuous improvement of its research methods and tools, the CPI 2012 will be calculated using an updated methodology, which will make it easier to trace how raw scores from the data sources are recalibrated for inclusion in the Index and will enable CPI scores to better capture changes over time from 2012 onwards.

Model – The Centre for Global Development's and the Brookings Institution's Quality of Official Development Assistance (QuODA) Index²⁶

The Centre for Global Development (CGD) and the Brookings Institution (Brookings) first published the Quality of Official Development Assistance (QuODA) Index in 2010. Its purpose was to complement the aid monitoring that was taking place through the Aid Effectiveness process with an independent quantitative assessment. It was intended to fill a data gap by being the only index that looked at more than one hundred agencies. The institutions spent two years developing their partnership, engaging in consultations with stakeholders and developing the measure before it was officially launched. They developed a measure, using publicly available data, that assesses donors across four dimensions, rather than providing an overall ranking. The four dimensions are intended to give donors specific enough information about their strengths and weaknesses so that they will know what actions they need to take to improve their practices.

QuODA has been aided by the fact that it was developed by two organizations that are well known, respected for their scholarship, and well regarded for their independence. It was the brainchild of CGD President Nancy Birdsall and Brookings Senior Fellow and Deputy Director, Global Economy and Development, Development Assistance and Governance Initiative, Homi Kharas, each of whom also is well regarded in the field of international development. Additionally, QuODA has received some media attention, which has helped raise its profile. CGD and Brookings have also presented QuODA publicly and to some of the ranked agencies at their request.

According to CGD, many donors have expressed interest in QuODA since its publication. Some are seeking to bolster their reputations. Some are using it as part of their reviews of their own practices. Some bilateral donors are looking for additional data to justify their support of multilateral agencies. Many are critiquing the data. CGD notes that, QuODA relies on data from the OECD because it is the most comprehensive source of standardised data, although the authors acknowledge its limitations. However, the data are a few years old by the time they are available, which limits QuODA's short-term utility. CGD and Brookings are drawing attention to this data gap and are using this as an opportunity to encourage government compliance with the International Aid Transparency Initiative (IATI), to which

²⁶ Interview with Rita Perakis, Center for Global Development, March 30, 2012.

many have signed on. Thus, QuODA is pushing for improved practices and improved availability of quality data.

CGD and Brookings wanted to include information about evaluation practices, in addition to aid delivery practices. However, they did not receive sufficient responses to a survey about evaluation practices sent to donor agencies. They also found that evaluation data were largely not publicly available, and where they were available, they were not comparable. They found that they could not quantitatively capture an indicator related to using evaluation results or the quality of evaluations. Instead, they settled on an indicator on evaluation policies, since this was the only available and comparable data point they could find. They developed the indicator by agreeing on a small number of key principles essential to include in an evaluation policy, based on OECD standards for evaluation practices, and then going to each agency's website to check its policy against that list. They have also highlighted to the OECD the need for an international reporting standard on evaluation practice and use.

Model – The World Bank's and the IFC's Doing Business Index²⁷

In 2002, the World Bank and the International Finance Corporation (IFC) launched the Doing Business Project as a way to help governments promote growth and create opportunities for the poor by developing their private sectors. The project goal was "to provide an objective basis for understanding and improving the regulatory environment for business around the world."²⁸ The project team decided to focus their efforts on gathering and analysing quantitative data related to business regulations and their enforcement for domestic small and medium-sized companies. In 2003, the team published its first report, which covered five business regulation topic areas, each represented by an indicator set, in 133 countries. By 2012, the report expanded to include 11 business regulation topic areas in 183 countries.

Prior to the launch of the Doing Business Project, only perceptions-based data were available. By creating these quantitative data sets, the project team allowed for comparisons of business regulation environments across countries and over time. While the data focus on business regulation environments in countries' largest business cities, subnational reports in select countries capture business regulation environments in additional cities and regions.

The original Doing Business indicators were initially developed based on talking with enterprises about the biggest constraints they faced. To limit the indicators included in the index, the Doing Business team narrowly defined its focus: business regulations and enforcement affecting small and medium-sized enterprises. They then partnered with academics who wrote background papers for each indicator, analysing their importance. These papers, which required multi-year research, have been or are being published in peer reviewed literature.

²⁷ Interview with Carolin Geginat, Sr. Private Sector Development Specialist, Global Indicators and Analysis Group (GIADB), Financial and Private Sector Development, The World Bank/IFC, August 2, 2012; Doing Business website, www.doingbusiness.org, accessed August 3, 2012.

²⁸ "About Doing Business," *Doing Business: Measuring Business Regulations*, the International Finance Corporation and the World Bank, <http://www.doingbusiness.org/about-us>, accessed August 3, 2012.

It has taken a few years to add each new indicator to the index. In the most recent example, the Doing Business team piloted the getting electricity indicators in 40 countries over a period of 10 months. This was followed by gathering the full data set. Then, the team produced the background paper for the set of indicators in house and presented it at research conferences for vetting purposes. A year later, the team included the indicators in the annex to the Doing Business Report, where it remained for two years. In the fourth year, after having fully tested the indicators and addressed stakeholder concerns, the team was able to place the indicators in the ranking.²⁹

Collecting data for the Doing Business indicators requires primary data collection. To gather these data each year, around 40 Doing Business team members work with more than 9,000 professionals and government officials who are experts on the Doing Business topics and who live in the countries covered by the Doing Business index. Survey respondents are asked to reference laws and regulations that inform their responses. Those laws and regulations are then checked to ensure the accuracy of the information being received.

The Doing Business team believed that a ranking could spark competition among the governments that it covered, since governments do not like seeing themselves ranked low or falling in a ranking. By producing press releases and working with journalists around the world to cover the Doing Business reports, they called attention to the data. The fact that the project is part of the World Bank helped to raise its profile and give it credibility.

However, the project team also noted that, because they are ranking their donors, along with their recipients, they must ensure that they are managing their relationships with care. They worked with their donors to get early buy-in to the project. Articulating a clear purpose from the beginning helped establish their donors' expectations. They also discussed the importance of basing the measures on independent analysis. Holding ongoing conversations as the team is developing new measures or revising current ones and prioritizing donor buy-in, even if it affects project timelines, have been critical for maintaining donor engagement.

The data can be summarized in a simple ranking number, which is helpful for gaining press coverage and initial government attention. It can also be disaggregated down to processes within a business regulation topic, so that governments can better understand their strengths and where reforms might be helpful. Because all data are publicly available, governments can identify other governments that might have experience that can inform their reforms. At the same time, the project team recognizes that the Doing Business data point to where change needs to happen and help identify opportunities for peer learning, but leave it to governments to convince their constituencies that change will be beneficial.

The project team holds some discussions with board members after the annual publication of the report, particularly about reforms that governments are advancing. Much dialogue between the project team and governments is initiated by the governments. Some express concern about the measures and the methodology. Others who are interested in pursuing reforms are seeking more information to help guide their processes.

²⁹ Interview with Carolin Geginat.

The Doing Business team is aware of its findings stimulating policy debates in at least 80 countries. It also knows that its publicly available data are being used by researchers to examine the linkages between company-level regulations and economic outcomes.³⁰ Around 30 governments have formed regulatory reform committees that report directly to the president or are at the inter-ministerial level and that include the Doing Business indicators among the inputs that help guide their programmes aimed at improving their business environments. Governments have informed the Doing Business team of more than 300 regulatory reforms they have implemented with input from the Doing Business index.³¹

Model – Publish What You Fund’s Aid Transparency Index³²

Publish What You Fund developed the Aid Transparency Index as a tool to capture and influence the practices of their advocacy targets, in part directly through the index and in part through the media attention it would generate. To create this index, in 2010, they examined publicly available secondary data sources on aid transparency, and found that the data were insufficient, out of date and hard to compare because of data gaps. As a result, they developed their own data collection instrument, comprised of a survey and two other data sources. For the survey, Publish What You Fund asks national NGO platforms to assess donors and then asks those donors to respond to the data presented, verifying and correcting as necessary. The 2012 Index contains 43 indicators grouped into three levels – organization, country and activity. Each level is weighted equally at 33%.

Publish What You Fund notes that indices can serve more like sticks than carrots for those being assessed. They have developed a communication strategy to mitigate this. They suggest engaging with agencies that will be assessed early on, requesting their input, preparing individual briefs for them, and inviting their feedback on the data and methodology before the ranking is published. In the first year of piloting the methodology, they gave some donors advance warning of the results but they do not anticipate doing this in future years. They have invested time building their reputation as go-to experts on aid transparency. They work closely with their four primary donor targets to better understand their challenges and make tailored recommendations to address those challenges. They also recognize donors who do well on the index with positive press releases.

Publish What You Fund’s relationships with their target donors range from very close to distant, as would be expected. They regularly conduct power mapping exercises to determine which donors might influence change in a positive way, which might block change, and which are somewhere in between. They focus their communication efforts accordingly. Depending on the circumstances and the donor’s response, they balance tough press releases and public statements with direct discussions. They actively encourage donors to reach out to them, if they are looking for ideas or feedback. In their more indirect advocacy, they also reach out to and brief other civil society organizations and

³⁰ *World Bank Report Finds More Economies Implemented Business Reforms in 2010-2011*, press release, Washington D.C., October 20, 2011.

³¹ *Doing Business 2012: Doing Business in a More Transparent World*, Washington, D.C.: The International Bank for Reconstruction and Development / The World Bank, 2012, 22.

³² Interview with Rachel Rank, Publish What You Fund, April 12, 2012; e-mail from Rachel Rank, July 25, 2012.

government representatives on relevant issues and help them with their messaging. This is part of their work to build a community of interest around aid transparency and encourage different organisations and voices to engage with the issue at a country or donor-specific level.

Publish What You Fund also pays significant attention to how it publicizes the Aid Transparency Index. It has developed an interactive website where visitors can play with and reweight the data.³³ In the report itself, it has included all the technical information in an annex, with the report body focusing on the ranking and providing information about each donor. Publish What You Fund suggests targeting different publicity opportunities for different donors: some might appreciate a press release or an op-ed; others might like a roundtable at which the findings can be discussed.

Model –CGAP’s SmartAid for Microfinance Index³⁴

In 2006, the Consultative Group to Assist the Poor (CGAP) brought together the heads of 29 major development institutions and CGAP members at the “Better Aid for Access to Finance” meeting. There, building on previous commitments and many years of working together, the participants signed the *Compact for Better Aid for Access to Finance*. One of the agreements within this compact was a request that CGAP develop an index to evaluate their internal management systems and, thus, “help [the participating] agencies assess and benchmark performance, provide incentives to make improvements, and further stimulate the international debate on the quality of aid.”³⁵

CGAP’s and participating members’ assumption is that having good internal management systems is a prerequisite for producing or contributing to effective development programmes. CGAP found that, when examining performance, the explanation for constraints or incentives is often related to internal management processes. While this is an early link within the development chain, it is one over which donors have the most control and, all acknowledge, where there is still room for improvement.

The SmartAid indicators emerge from the peer reviews that CGAP led in 2002 and 2003. The peer reviews involved teams of CGAP staff and member agencies evaluating participating organizations’ internal systems, and led to significant learning regarding what is effective and what is not. While time consuming, the peer reviews built relationships and ensured that organizational systems were thoroughly assessed. This was very highly valued by participating organizations.

SmartAid depends on examining primarily internal documents that participating organizations submit. Over time, the SmartAid team has learned what documents organizations have, which are most useful, and how to ask for them so that they receive only those that are most relevant. SmartAid has an online platform where members can

³³ See <http://www.publishwhatyoufund.org/resources/index/2011-index/>.

³⁴ Interview with Barbara Scola-Gähwiler, Donors and Investors Team, SmartAid for Microfinance Index, CGAP, July 24, 2012; and CGAP’s SmartAid for Microfinance Index webpage, <http://www.cgap.org/p/site/c/template.rc/1.11.7956/1.26.3224/>, accessed July 26, 2012.

³⁵ CGAP, *Compact for Better Aid for Access to Finance*, Paris, France, October 20, 2006, p. 2, http://www.cgap.org/gm/document-1.9.2152/BetterAid06_Compact.pdf, accessed July 26, 2012.

upload documents and CGAP can easily access them. The platform automatically assigns each document a number, which helps with tracking information. Because participating organizations are CGAP members, they trust CGAP and are willing to spend significant staff time looking for the documents. Many turn this data collection exercise into their own internal learning process.

SmartAid depends on examining primarily internal documents that participating organizations submit. Over time, the SmartAid team has learned what documents organizations have, which are most useful, and how to ask for them so that they receive only those that are most relevant. SmartAid has an online platform where members can upload documents and CGAP can easily access them. The platform automatically assigns each document a number, which helps with tracking information. Because participating organizations are CGAP members, they trust CGAP and are willing to spend five days looking for the documents. Many turn this data collection exercise into their own internal learning process.

SmartAid's data collection approach has strengths and weaknesses. It risks missing aspects of participants' systems that are not captured in documents. For example, in the pilot round, CGAP had an indicator on in-country collaboration: very important to effectiveness, but not captured in documents, and therefore, not included going forward. On the positive side, the SmartAid indicators provide a data collection approach that is much more sustainable than the peer reviews, given the lower level of effort required. CGAP believes that this approach is justified, since capturing organizational information in writing is important to organizational effectiveness.

During the biennial assessment period, two CGAP staff members allocate up to half of their time or less to the assessment, and four external peer reviewers dedicate approximately 25 days each. With this investment, CGAP believes it could assess up to 20 organizations in one round, although the maximum number of participating organizations to date has been 11. While this represents a smaller level of effort than that which was required for the peer reviews, it is not inconsequential.

In developing the index, CGAP consulted with members, consultants, the head of OECD DAC and CGD, among others. CGAP also reviewed standardized microfinance measures. Out of its peer review experience, the review of the measures and the consultations, CGAP developed an initial set of 20 indicators and piloted them with seven member organizations in 2007. Based on the pilot, CGAP was able to narrow the index to nine indicators. Eleven member organizations then joined the 2009 round, which served as a baseline for a biennial assessment.

Participating organizations receive a report detailing strengths and weaknesses, highlighting good practices, offering recommendations and providing an overall score. Throughout the process, SmartAid staff members engage in discussion with participants and provide technical assistance, as necessary. By highlighting good practices, the SmartAid team identifies opportunities for peer learning that CGAP sometimes facilitates.

CGAP members describe a number of benefits of participating in SmartAid. Primarily, SmartAid serves as external validation for the changes that member organizations have

already identified, while also highlighting needed changes they might have overlooked. It helps organizations prioritize these changes and encourages management to act. Gathering the documents required for participation in the assessment can serve as an internal stocktaking exercise. It can be helpful for capturing institutional history, reflecting on organizational processes and practices, and orienting new staff members. The index can also let participants know how they compare to other participating organizations, helping them connect with others who have complementary strengths. According to participating organizations, SmartAid is a relatively low effort way to get useful advice. Their participation is seen as evidence of their commitment to transparency and accountability.

Yet, only some participating CGAP members use SmartAid as a tool to help them improve their systems. CGAP notes that organizations that rank in the middle tend to respond to the incentive, asking CGAP for concrete recommendations and sometimes technical assistance. For poorly performing organizations, SmartAid has proven less effective, as a low score sometimes triggers a defensive reaction rather than proactive implementation of the recommendations. CGAP engages with participating organizations after the assessment to help them address their weaknesses. However, given the resources involved, CGAP must do this selectively.

Because participating organizations are CGAP members, CGAP makes their relationship its highest priority and has made a number of choices in this regard. To help protect its relationships with members, as well as to add credibility to the assessment, SmartAid scores are assigned by an independent review board. CGAP does not make it mandatory for agencies to publish their reports. However, it encourages participating agencies to do so. Also, CGAP does not seek out media attention for SmartAid.

In addition, CGAP has not pushed for agencies to participate in every biennial assessment. Rather, CGAP wants organizations to participate because it will contribute to their internal change processes. CGAP also wants to give organizations that have participated in the past sufficient time to respond to recommendations they have received. The changes made in response to SmartAid should be reflected in an organization's future SmartAid scores. SmartAid's aim is to spur dialogue about effectiveness. In SmartAid, CGAP prioritizes learning over accountability.

CGAP is only committed to continuing SmartAid as long as CGAP and its members find it useful. CGAP has thought about outsourcing SmartAid to a consultant, but members said it is important to have CGAP, with its stature and credibility, leading this. CGAP has tested a self-assessment as a lighter version. Participants got to less complicated lessons quickly, but were less able to identify and address complicated elements. CGAP believes that some combination of self-assessment and CGAP intervention might be an effective way to achieve SmartAid's objectives with a lower level effort and is exploring this possibility.

CGAP has also considered leaving the index to focus on peer learning around key issues, which is a very important aspect of SmartAid. CGAP had an event in July 2011 bringing together SmartAid participants and others in a two-day workshop. In the second day, they focused on specific issues touched on in SmartAid. Different agencies talked about how they addressed these issues, followed by discussion. The SmartAid indicators related to

accountability, on which participating organizations consistently score low, would lend themselves to future peer learning events.

SmartAid is one of CGAP's tools for helping members improve their practices. While only a subset of CGAP's members have participated directly in SmartAid, all members hear about SmartAid conversations. CGAP also recommends portfolio reviews, which can help an organization learn why things work well in some places but not in others, and why some things work and others do not. Throughout the year, CGAP engages with members on issues in microfinance in such settings as meetings, such as their annual meeting, and through a variety of peer learning opportunities.

Annex 3. Interviews

Name	Organization
Rosalind Eyben	Big Push Forward
Rita Perakis	Centre for Global Development (CGD)
Laura Loo	China Agriculture University, Beijing
Barbara Scola-Gähwiler	Consultative Group to Assist the Poor (CGAP)
Gretchen Shanks	Bill and Melinda Gates Foundation
Orlando Gracia	Government of Colombia
Ole Andersen	Government of Denmark (DANIDA)
Sulley Gariba	Government of Ghana
Gonzalo Hernández L.	Government of Mexico (CONEVAL)
Marie Gaarder	Government of Norway (NORAD)
Ian Goldman	Government of South Africa
Antoine De Kemp	Government of the Netherlands (MOFA/IOE)
David Bonbright	Keystone Accountability
Hans Lundgren	OECD
Albert Byamugisha	Office of the Prime Minister, Government of Uganda
David Smith	Office of the Prime Minister, Government of Uganda
Rachel Rank	Publish What You Fund
Penny Hawkins	Rockefeller Foundation
Peter Eigen	Transparency International
Deborah Hardoon	Transparency International
Nick York	U.K. Government (DfID)
Indran Naidoo	UNDP
Caroline Heider	World Bank
Barbro Hexeberg	World Bank

Nidhi Khattri	World Bank
Naoko Watanabe	World Bank
Carolyn Geginat	World Bank Doing Business Index
Barbara Gahlwiler and Zekebweliwai Geh	CGAP SmartAid for Microfinance Index

Bibliography

Busan Partnership for Effective Development Co-operation, Fourth High Level Forum on Aid Effectiveness, Busan, Republic of Korea, November 29 - December 1, 2011, December 1, 2011,

http://www.aideffectiveness.org/busanhl4/images/stories/hlf4/OUTCOME_DOCUMENT_-_FINAL_EN.pdf, accessed May 22, 2012.

Carden F., *Knowledge to Policy: Making the Most of Development Research*, IDRC, Ottawa, Canada, 2009.

CONEVAL, Creación del Consejo Nacional de Evaluación de la Política de Desarrollo Social, <http://www.coneval.gob.mx/cmsconeval/rw/pages/quienessomos/creacion.es.do>, accessed July 27, 2012.

Consultative Group to Assist the Poor (CGAP), *Compact for Better Aid for Access to Finance*, Paris, France, October 20, 2006, http://www.cgap.org/gm/document-1.9.2152/BetterAid06_Compact.pdf, accessed July 27, 2012.

Consultative Group to Assist the Poor (CGAP), *Improving Effectiveness from Within: SmartAid for Microfinance Index*, Brief, Paris, France, December 2009, http://www.cgap.org/gm/document-1.9.40934/BR_SmartAid_Microfinance_Index.pdf, accessed July 27, 2012.

Consultative Group to Assist the Poor (CGAP), *SmartAid for Microfinance Index*, <http://www.cgap.org/p/site/c/template.rc/1.11.7956/1.26.3224/>, accessed July 26, 2012.

Gaarder M. and Briceño B. *Institutionalisation of Government Evaluation: Balancing Trade-Offs*, Working Paper No. 8, 3ie, New Delhi, July 2010.

International Aid Transparency Initiative, <http://www.aidtransparency.net/>, accessed July 27, 2012.

International Initiative for Impact Evaluation (3ie), "The Commitment to Evaluation (c2e) Indicator: A Discussion Note," New Delhi, India, January 2012, unpublished.

Keystone Accountability, *Keystone Performance Surveys*, <http://www.keystoneaccountability.org/services/surveys>, accessed June 17, 2012.

Leeuw F., Rist R. and Sonnichsen R., *Can Governments Learn? Comparative Perspectives on Evaluation and Organizational Learning*, Transaction Publishers, New Brunswick, New Jersey, 1994.

MacKay K., *How to Build M&E Systems to Support Better Government*, World Bank, Washington DC, 2007.

MacKay K., *Institutionalization of Monitoring and Evaluation Systems to Improve Public Sector Management*, ECD Working Paper Series - No. 15, World Bank, Washington, DC, January 2006.

National Evaluation Capacities: Proceedings from the 2nd International Conference, 12-14 September 2011, Johannesburg, South Africa, Evaluation Office, United Nations Development Programme, March 2012.

Open Government Partnership, <http://www.opengovpartnership.org/>, accessed July 27, 2012.

Peer Review: An OECD Tool for Co-operation and Change, OECD, Paris, 2003, <http://browse.oecdbookshop.org/oecd/pdfs/free/0103083e.pdf>, accessed May 28, 2012; interview with Hans Lundgren, OECD, April 2, 2012.

Publish What You Fund, *2011 Pilot Aid Transparency Index*, <http://www.publishwhatyoufund.org/resources/index/2011-index/>, accessed July 27, 2012.

Roodman D., *Building and Running an Effective Policy Index: Lessons from the Commitment to Development Index*, Centre for Global Development, Washington DC, March 2006.

Székely M., *Promoting Commitment to Evaluate* (New Delhi, India: International Initiative for Impact Evaluation, May 2012).

Székely M., "The use of evidence in Mexico: experiences and challenges," *Journal of Development Effectiveness*, Volume 3, No. 1, March 2011.

Székely M., "Toward results based social policy design and implementation", Centre for Global Development, Washington DC, February 2011.

Transparency International, <http://www.transparency.org/research>, accessed June 20, 2012.

White H., Executive Director, International Initiative for Impact Evaluation (3ie), conversation with Christelle Chapoy, shared in an e-mail to Carlisle Levine, May 14, 2012.

Publications in the 3ie Working Paper series

Building on what works: commitment to evaluation (c2e) indicator by Carlisle J. Levine and Christelle Chapoy, Working Paper 18, February 2013

From impact evaluations to paradigm shift: A case study of the Buenos Aires Ciudadanía Porteña conditional cash transfer programme by Gabriela Agosto, Eduardo Nuñez, Horacio Citarroni, Irma Briasco and Nicolás Garcette, Working Paper 17, August 2012

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 by Gala Díaz Langou and Paula Forteza, Working Paper 16, August 2012

Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework by Howard White and Daniel Phillips, Working Paper 15, July 2012

Behind the scenes: managing and conducting large scale impact evaluations in Colombia by Bertha Briceño, Laura Cuesta and Orazio Attanasio, Working Paper 14, December 2011

Can we obtain the required rigour without randomisation? by Karl Hughes and Claire Hutchings, Working Paper 13, August 2011

Sound expectations: from impact evaluations to policy change by Vanesa Weyrauch and Gala Díaz Langou, 3ie Working Paper 12, April 2011

A can of worms? Implications of rigorous impact evaluations for development agencies by Eric Roetman, 3ie Working Paper 11, March 2011

Conducting influential impact evaluations in China: the experience of the Rural Education Action Project by Mathew Boswell, Scott Rozelle, Linxiu Zhang, Chengfang Liu, Renfu Luo, Yaojiang Shi, 3ie Working Paper 10, February 2011

An introduction to the use of randomized control trials to evaluate development interventions by Howard White, 3ie Working Paper 9, February 2011

Institutionalisation of government evaluation: balancing trade-Offs by Marie Gaarder and Bertha Briceno, 3ie Working Paper 8, July 2010

Impact Evaluation and interventions to address climate change: a scoping study by Martin Prowse and Birte Snilstveit, 3ie Working Paper 7, March 2010

A checklist for the reporting of randomized control trials of social and economic policy interventions in developing countries by Ron Bose, 3ie working paper 6, January 2010

Impact evaluation in the post-disaster setting by Alison Bутtenheim, 3ie Working Paper 5, December 2009

Designing impact evaluations: different perspectives, contributions from Robert Chambers, Dean Karlan, Martin Ravallion, and Patricia Rogers, 3ie Working Paper 4, July 2009. Also available in Spanish, French and Chinese

Theory-based impact evaluation by Howard White, 3ie Working Paper 3, June 2009. Also available in French and Chinese.

Better evidence for a better world edited by Mark W. Lipsey University and Eamonn Noonan, 3ie & The Campbell Collaboration, 3ie Working Paper 2, April 2009

Some reflections on current debates in impact evaluation by Howard White, 3ie Working Paper 1, April 2009

For the latest 3ie working papers visit:

http://www.3ieimpact.org/3ie_working_papers.html