

Terms of Reference Research Consultancy –



White paper for the Commitment to Evaluation Indicator International Initiative for Impact Evaluation (3ie)

Background

Experience to date shows that the use of evidence by donors and governments when designing and adopting development programmes remains sporadic. There are many examples where a programme was shown to have no impact but was expanded, as well as examples of programmes with positive impact being terminated. To promote better use of evaluation evidence in policy making and programme design, 3ie is launching a Commitment to Evaluation (c2e) indicator. The indicator will provide a measurement of government and donor agency use of evaluation evidence allowing for recognition and reward for progress and good practice. The indicator will be developed and piloted in 2012 for donor agencies with the intent to recognize donors that make systematic use of evidence and thus motivate others to do the same.

3ie's initiative follows the example of other successful efforts to use awards or indexes to focus the attention of policymakers. Indexes such as the UN Development Programme's Human Development index, Transparency International's Corruption Perception index, and the Centre for Global Development's Quality of ODA (QuODA) index have raised awareness on key issues and influenced practice of governments and development agencies. The Mexican National Council for the Evaluation of Social Development Policy (CONEVAL) annual award for good practices in social evaluation has strengthened political buy-in and commitment to evaluation in Mexico. In developing this c2e indicator, 3ie will draw from the lessons learned by similar initiatives on how best to motivate and award evaluation practices and build and run an effective cross-agency and cross-country indicator. More detailed background information on the rationale and theory of change behind the project is available in the discussion note in the annex.

3ie has headquarters offices in New Delhi, a Systematic Review Office in London, and a liaison office in Washington DC. We are seeking a senior researcher to be contracted as an independent consultant to write a white paper in two parts that will develop and test a theory of change for the indicator, propose details for constructing the measure, and a process for collecting the necessary data.

Scope of Work

The researcher will work under an independent consultancy contract. The tasks will include:

- Review a selection of current cross-agency and cross-country indicators, distilling the lessons learned from the experience of these indicators, the choice of indicators to be reviewed will be agreed with 3ie;
- Further develop the theory of change and framework for the index;
- Research and map existing data that may be used for calculating a c2e indicator;
- Interview evaluation units and departments within donor agencies and governments to gain further insights on some of the existing incentives, bottlenecks and evaluation cultures within governments, donors and multilaterals;
- Make presentations at venues to be agreed with 3ie;
- Co-author Part I of the white paper with 3ie staff which will look at some of the underlying assumptions in the theory of change and at how best to make the indicator work. Co-author Part II of the white paper with 3ie staff, which will propose a specific measure to be used as the indicator along with the methodology behind the proposed measure and the data collection process necessary to calculate that measure.

The deadline for Part I of the white paper will be eight weeks from the effective date of the consulting agreement. The consultant may bill up to 20 LOE days (160 hours) during that time period. The deadline for Part II of the white paper will be six weeks from the date that Part I is submitted, and the consultant may bill up to 15 additional LOE days (120 hours) during that period. The consultant will report to the Senior Advocacy and Policy Officer.

Qualifications

Desired qualifications include:

- Experience in development evaluation, preferably with working knowledge of experimental and quasi-experimental methods of rigorous impact evaluation;
- Knowledge of international development policies and programmes;
- Experience working with donor agencies and governments in low- and middle-income countries;
- Strong analytical writing skills in English.
- PhD degree or Masters degree with demonstrated research experience in economics, political science, or other relevant social science;

Applications

Please submit a cover letter stating how the work will be approached, CV(s), and writing sample to cchapoy@3ieimpact.org by 29 February, 11.55 pm Indian Standard Time.

.Only finalists will be contacted. You may also send any questions to ChristelleChapoy.

Annex

The Commitment to Evaluation (c2e) Indicator A discussion note

About the c2e indicator

The Commitment to Evaluation Indicator is an annual indicator of the production and use of evaluative evidence by development agencies and governments in low and middle income countries.

Published annually, c2e will recognize donor agencies' and governments' efforts to base policy on evidence, providing an incentive for others to do the same. The indicator sends the signal that development spending should be guided by evidence in order to maximize its impact in improving people's lives.

This discussion notes lays out our thinking on c2e, ending with a list of questions for discussion as we take the idea forward.

Why have a Commitment to Evaluation indicator?

Indices such as the UNDP's Human Development index, Transparency International's Corruption Perception index, the Centre for Global Development's Quality of ODA (QUODA) index have been successful in raising awareness on key issues, and so influenced development practice. The National Council for the Evaluation of Social Development Policy (CONEVAL) annual award for good practices in social evaluation in Mexico has strengthened political buy-in and commitment to evaluation.

Similarly, awarding evaluation practice will raise awareness amongst policymakers and programme managers that they can take credit for initiating rigorous evaluation studies. Specifically, c2e will:

- Provide an incentive for governments and donors to use evidence from evaluations
- Promote the idea of evaluation as a good thing rather a threat from the danger of adverse findings
- Encourage governments and donors to identify success and learn from failures

Realization of these goals will increase the use of evidence, and so promote effective development.

How will c2e work?

3ie will collect data to produce and publish c2e, and then announce the awards for good practice. For these activities to succeed in the ultimate goal of increasing development effectiveness there are key assumptions in the underlying theory of change which need to be satisfied.

The key assumptions are that:

1. People are aware of the indicator
2. These people care about their country or agency's performance in the indicator
3. These people are in a position to influence the production and use of evidence in their country or agency
4. That 'doing well' on the indicator does indeed bring about policy changes which promote more effective development

3ie can address the first and third of these assumptions through targeted dissemination. The second is addressed partly through advocacy on the importance of evaluation. We also hope to draw on competition amongst donor agencies in wanting to perform well against each other and countries' desire to not to be seen to be performing poorly. Donors will be encouraged to use the indicator in their own assessment of the quality of country policies, which will provide an additional incentive to countries.

The final assumption is part of the rationale for 3ie. But we recognize the need to produce more evidence of the difference good evaluation can make.

Questions for discussion

We would appreciate feedback on the following issues:

- Should the indicator be mainly focused on impact evaluation or should it cover evaluation more broadly?
- Given the answer to the previous question, what is the appropriate name for the indicator or scorecard?
- How do we capture context-specific good practice in the use of evaluation?
- An index, scorecard, or other type of measure?
- What sources of data can be used? If surveying organizations, who should we survey? Should we use multiple surveys for a single organization?

Next steps

1. Informal consultation
2. Commission a background paper to develop and test the theory of change of the indicator. This will involve looking more specifically at: Do rankings matter? Do awards matter? Do scores matter? Do our target agents only compare themselves to themselves (or to some external standard) or are they more motivated by being compared to others? What information does the indicator need to include to induce behavior change?
3. Establish a working group or steering group (with written TOR)
4. Plan a consultation with the OECD DAC Evaluation Network
5. Develop a framework for the index and award

6. Draft a white paper with methodology for the index and award
7. Test and pilot a developing country survey
8. Plan a consultation at the NONIE meeting
9. Collect data
10. Analyse the data and estimate the indicators for each donors and government
11. Launch the first awards

Annex 1: Summary report of consultation survey

3ie conducted a consultation survey around the Busan meeting on Aid Effectiveness to assess whether people assimilate a commitment to result with a more systematic use of evidence in policymaking.

The survey consultation was widely disseminated amongst development practitioners, donors, developing country policymakers and evaluators through various listservs, communities of practice and targeted mailings. The consultation lasted over a period of two weeks and we received a very good response showing a strong interest in the initiative with 220 completed survey.

What we have learned

Over 80% agreed that commitment to evaluation is very important for development effectiveness with 98% said it is important.

For over 64% of respondents, commitment to evaluation means using evidence to improve design and implementation of programmes and policies. For 19%, it is about having a strong culture of M&E and for over 15% it means using evaluation for scaling up or scaling down an intervention.

In terms of good measures for use of evidence, 83% identified the requirement to look at evidence when designing a programme as a good measure. For 81% publishing evaluation results was another good measure and 67% identified the practice of conducting an impact evaluation of a pilot programme.

On the question regarding how respondents perceived how their organization was communicating what has been learned from evaluation, the majority (69%) was scoring between average to good.

Some opposition to the Index:As part of the consultation we have received two negative feedback from individuals strongly opposed to having a credible index: *"There can be no standardized indicators or observeables. Commitment is situational and complex, not simplistic and mechanistic. 3ie is trying to simplify the complex which, of course, is the 3ie brand "Simplemindedness about complex evaluation issues." An index of evidence use is not only simple-minded and mechanistic but dangerous. Such an approach ignores all the evidence about the corruption of indicators in high stakes performance measurement systems."*

“This index is a poor idea, and presumably meant as a PR/propaganda tool - as the CGD index is. Measurements will be very heavily based on judgments and thus easy to manipulate within very broad ranges. The simplistic questions here reinforce that suspicion. Don't you have anything better to do?”

Some other comments were more stressing the need to make a clear distinction between “use of evaluation” and “use of evidence”; the need to be clear as per whom the index is targeted at “donors” versus “developing countries”; and the fact that the “push for quantifiable indicators is counter-productive” as the concept of commitment to evaluation “is primarily about organizational culture”.

Overall the suggested indicators and observable characteristics can be regrouped around the following categories*:

** Only specific quantifiable and observable characteristics from the answers provided by respondents are regrouped and synthesized below*

On practice: The number of rigorous impact evaluations (by independent evaluation institutions) that a policymaker has demanded during the last five years.

Percentage of programmes evaluated by an independent evaluation.

Spending on M&E; amount of resources going to the evaluation as % of total resources; or the budget to hire or conduct evaluations as a percentage of total cooperation budget or ODA

Degree of rigor in existing studies (e.g. Create a scale along "simple before-after" vs. "RCT")

Conducting and/or funding impact assessments with counterfactuals at pilot. If no pilot, then build impact assessment into program design.

Compliance with Evaluation Cooperation Group standards used as evaluation framework

Frequency of evaluation activities within a project or programme

Organisation evaluation culture: My organization has an evaluation policy or framework

My organization has established evaluation standards.

My organization periodically holds evaluation reviews to glean lessons learned.

My organisation supports/funds capacity building of executing/implementing agencies to design and conduct robust baseline and endline surveys.

My organisation works jointly with executing/implementing agencies in designing programme/project logframes.

Staff induction and job roles all include some involvement and responsibility for evaluation

Requirement that relevant previous evaluations be referenced when designing new programs -- with evidence that previous lessons learned are being applied.

Presence of a serious self-assessment system among staff Number of managers requiring evaluation results for making a decision

Organization has a written record of how it responded to evaluation findings/recommendations, that can be verified

Do the head of evaluation and other senior evaluation staff in the org know how to - and actually promote - the use of randomization, where possible, to evaluate their programs?

Does the organization have an independent evaluation office that reports directly to the board in the case of a private organisation or to the parliament in the case of a donor.

Financial investment in, and proof of implementation of, evaluation capacity training for staff at all levels.

Use of evidence: Number of scaled up projects due to evaluation findings

Number of suspended (or not extended) projects due to evaluation findings

Percentage of evaluation recommendations that are followed up on

Number of programmes (new or new phases of programmes) that are clearly designed on the basis of existing evaluation based knowledge - e.g. documented through direct or indirect references in programme documents to evaluation insights.

Project proposals have to refer to lessons learned from similar projects and quality assurance committee/investment committee verifies whether this has been included (and makes sense...)

Managers are accountable for developing a response to evaluation results and sharing with stakeholders specific steps they will take to implement this response.

Number of projects funded despite proof that they are ineffective (e.g. computers in classrooms, teacher training, ...)

My organization has follow-up mechanisms to the findings of evaluations

Dissemination/transparency: The sharing/publication of data and evaluation findings. % of evaluations whose results are announced and published within an organization

Evaluations are discussed with stakeholders / beneficiaries and agree together on next steps

Publication of the analysis of 'failed' programmes.

Dedicated (and protected) work time for group reflection of evaluative information pertaining to programming monitoring and evaluation collected data

ALL details get published: details of the programme, of the evaluation, details of sample sizes (for programme & evaluation & control group), how they were chosen (incl. precise method of randomising)... Good & bad data is published. This is science: so it all needs to be open.