

The 'gold standard' is not a silver bullet for evaluation



Harry Jones

'A focus on the institutional capacity and incentives that foster the use of evaluations is overdue, and is the key to promoting learning and accountability'

Increased aid flows, and a growing focus on the effectiveness of aid, has led to heightened interest in evaluating impact. Impact Evaluation (IE) examines the effects of development interventions beyond the direct work of the programmes to look at their contribution to people's wellbeing (Riddell, 2008).

IE should play two key roles: first, ensuring that learning from experience translates into more effective development programmes, and second, providing accountability to ensure that money is well spent. The current donor response to this challenge sees growing momentum to commission and produce IEs, with a particular emphasis on impact evaluations that involve 'experimental' or 'quasi-experimental' methods. These are seen, by some, as the only evaluations that are rigorous.

While the increased funding for IE is a step in the right direction, there needs to be a pluralistic approach to IE methodology: experimental and quasi experimental methods are not the only ones that are 'rigorous'. But more importantly, funding IEs and ensuring that they are carried out with rigorous methodologies are only two elements in a much wider set of practices that are needed to promote learning and accountability.

For one, it is crucial to foster the use of evaluations (Jones et al., 2009). This must, in turn, be buttressed by the appropriate institutional capacities and incentives (Foresti, 2007). A recent survey of IE databases found that nearly all those published contain favourable results (Jones et al., 2009). This is too good to be true, and an indication that there is still a long way to go to put in place solid processes for learning and accountability.

It is increasingly urgent that the donor community focuses on these issues. Seeing the challenge of evaluation as primarily methodological, and treating experimental and quasi-experimental IEs as a 'silver bullet' to improve the evaluation function could exacerbate the problem. Examining the current institutional incentives around IE use provides a warning that, without proper attention to such issues, certain practices around IE could risk creating

conflicting incentives that might even discourage learning and accountability. This warning was a strong theme in over 60 interviews conducted with evaluations experts and is echoed in available evidence on the use of IEs (ibid.).

Incentives for IE

There is preliminary evidence of two trends within agencies that threaten to skew incentives for informed policy and effective development programmes. First, many major donors (e.g. the World Bank) prefer particular types of IE that are seen as 'the gold standard' (EES, 2007). These are often based on 'counter-factual' methodology (comparing what has happened to what would have happened without the programme), assessed using experimental or quasi-experimental methods (ADB, 2006). Despite some claims to the contrary, proponents see this as the only objective way to evaluate interventions, while painting other methods as merely collecting 'opinions' (CGD, 2006). And, while mixed methods are discussed, this perceived hierarchy of evidence remains.

Second, these impact evaluations are commissioned mainly for upwards accountability and 'legitimation' purposes (Raitzer and Winkel, 2005; Jones et al., 2009). Proving to donors that an intervention has had some impact protects existing funding and boosts the chances of funding in the future. Equally, where projects, programmes and even whole sectors struggle to demonstrate impact, they may lose funding. Some institutions take this further, using experimental and quasi-experimental IE as a central pillar of results-based management approaches. On the surface, such IEs are ideal when allocating budgets to continue, modify or scale up interventions, and in conducting Cost-Benefit Analyses of the fall in poverty for every dollar spent.

While legitimation is worthwhile, and a key element of accountability, using evaluation to legitimise funding is only right when every intervention has a fair chance to demonstrate that it has had an impact. Experimental IEs are just one methodology among many. Like any other, they have strengths and weaknesses, suitable for some interventions and not others.

Overseas Development Institute

ODI is the UK's leading independent think tank on international development and humanitarian issues.

ODI Opinions are signed pieces by ODI researchers on current development and humanitarian topics.

This and other ODI Opinions are available from www.odi.org.uk

Table 1: The ‘gold standard’ versus other approaches

Characteristics of interventions favoured by ‘gold standard’ incentives	Characteristics of effective, accountable interventions
A focus on short-term results, simple and short causal chains, and on quantifiable effects that can be predicted in advance with confidence.	A focus on the long-term nature of sustainable change, the complexity and multidimensionality of problems and unexpected outcomes and effects (Ramalingam and Jones, 2008).
A need for interventions that involve standardised outputs and elements that are distributed and controlled from a central point.	A need for flexible interventions, responsive to context and unfolding events.
Aim to claim a specific portion of the effect of a particular intervention, usually a goal that aligns with donor priorities.	Incorporates Paris Declaration principles of partnership, harmonisation, and increased recognition for accountability to beneficiaries.

Learning and accountability

Experimental and quasi-experimental methodologies require a ‘dosing’ model, with interventions modelled as delivering a discrete and homogenous output, like the distribution of pills. They require a plausible counter-factual (finding a group large enough to represent a case similar to those receiving the intervention in all relevant variables). On these terms, it is easier for some sorts of programmes to demonstrate beneficial impact, and harder for others. It suits, for example, the provision of vaccines or school dinners, cash transfer-based social protection programmes or distribution of new seed varieties. Work in other areas is less amenable to such approaches. This includes research communication and advocacy, where many complex, interacting factors produce change (rather than any single programme), and sector-wide approaches, where it is impossible to identify a plausible comparison group to represent what would have happened without the intervention. These policy areas could come under unwarranted pressure, or lose funding.

This bias is not only unjustified, it could generate incentives that go against key practices and hard-learned lessons about aid effectiveness, accountability and learning, and how change happens (Table 1).

Results-based approaches to impact would strengthen this trend, further institutionalising the imbalance and endangering learning for development interventions. Individuals and organisations may be unwilling to admit error and learn from unexpected outcomes, undermining attempts to innovate,

and generating evidence that confirms institutional prejudices (Proudlock and Ramalingam, 2008).

The solutions

First, donors and those commissioning evaluation, need a more balanced view of ‘rigour’ and ‘evidence’. Experimental methodologies are only one way of looking at the impact of an intervention, and other methodologies can be just as rigorous and objective. Evaluation of a counter-factual is only one way to look at causality, and is applicable to less than 25% of policy areas (Jones et al., 2009).

Some argue that, in the social sciences, the best way to understand cause and effect is to look at why people change their behaviour. Other approaches emphasise the importance of looking at ‘causal packages’, the configurations of factors that cause outcomes (EES, 2007). This involves work from a range of disciplines beyond econometrics. Examples of alternative methodologies include: outcome mapping; utilisation-focused evaluation; most significant change; and ‘realistic’ evaluation.

Second, more attention must be paid to the institutional factors and incentives around accountability and learning (see Jones et al., 2009; Foresti, 2007; Proudlock and Ramalingam, 2008). Results-based management approaches to impact may be inappropriate in development because, for many projects, the final effect on people’s welfare cannot be predicted. Donors should instead foster communication of results and engagement between evaluators and decision-makers. Increased funding should go to strengthening capacity and organisational processes, linking impact evaluation to decision making.

Work is needed to maximise the influence of experimental and quasi-experimental IEs on improving practice. It is important to coordinate studies and conduct synthesis reviews, and donors should help strengthen the evidence base by publishing failures as well as successes. It is crucial to recognise that experimental and quasi-experimental IEs are just one method, and that, despite promises of demonstrating impact in line with agency goals, this is not always possible.

By Harry Jones, ODI Research Officer (h.jones@odi.org.uk). For further resources visit www.odi.org.uk/RAPID



Overseas Development Institute

111 Westminster Bridge Road, London SE1 7JD

Tel +44 (0)20 7922 0300

Fax +44 (0)20 7922 0399

Email publications@odi.org.uk

Readers are encouraged to quote or reproduce material from ODI Opinions for their own publications, but as copyright holder, ODI requests due acknowledgement and a copy of the publication.

© Overseas Development Institute 2009

ISSN 1756-7629

References

Asian Development Bank (ADB) (2006) *Impact evaluation: Methodological and operational issues*. Manila: ADB.

CGD (2006) ‘When Will We Ever Learn? Improving Lives through Impact Evaluation’. Report of the Evaluation Gap Working Group. Washington, DC: CGD.

European Evaluation Society (EES) (2007) ‘EES Statement: The importance of a methodologically diverse approach to impact evaluation – specifically with respect to development aid and development interventions’. Nijnerk: EES Secretariat.

Foresti, M. (2007) ‘A Comparative Study of Evaluation Policies and Practices in Development Agencies’. Report for AFD Evaluation Department.

Jones, N., Steer, L., Jones, H. and Datta, A. (2009) *Improving Impact Evaluation Production and Use*. ODI Working Paper 300. London: ODI.

Proudlock, K. and Ramalingam, B. (2008) ‘Re-thinking the Impact of Humanitarian Aid: Background Paper for the 24th ALNAP Biannual’. London: ALNAP.

Raitzer, D. and Winkel, K. (2005) ‘Donor Demands and Uses for Evidence of Research Impact – the Case of the Consultative Group on International Agricultural Research (CGIAR)’, prepared on behalf of the CGIAR Standing Panel on Impact Assessment.

Ramalingam, B. and Jones, H. with Young, J. and Reba, T. (2008) *Exploring the Science of complexity: ideas and implications for development and humanitarian efforts*. ODI Working Paper 285. London: ODI.

Riddell, R. (2008) ‘Measuring impact: the global and Irish aid programme’. Final report to the advisory board for Irish Aid.