



Evidence for whom?

3ie Washington Evidence Week Conference, 27 April 2017

Edited transcript of closing remarks by Richard Manning, chair, 3ie Board of Commissioners

I am not an evaluator – I am, I suppose, a development practitioner interested in improving the effectiveness of policies and programmes that contribute to development.

In a way, it shouldn't matter that my career interest has been in development. Any public servant – or, indeed, any interested taxpayer, citizen or civil society organisation – should be interested in the effectiveness of policies and programmes agreed through democratic or other acceptable objective-setting processes.

It is fairly obvious that a lot more rigorous scrutiny is needed of important public programmes in almost every country in the world. We have heard from the previous panel that the production of evidence is not a guarantee of reform, but it is surely one important element in leading to improvements over time. Public funds are a scarce resource, and their use needs objective scrutiny. I don't apologise for emphasising this point today in this city [Washington, DC].

However, being involved in 'development' does involve three particular issues:

1. Our sector has done more evaluations of donor-supported programmes than you will find for pretty well any other form of public expenditure – raising rather forcefully the question of what we have managed to learn.
2. We may be inclined to evade evaluation judgements in a stream of finance that is itself frequently contested (a reason why CGD initiated the *When will we ever learn?* report¹).
3. Most programmes we finance are carried out by governments or private institutions whom we do not control, but who have a lot less power and resources than we have to evaluate these programmes or to critically assess the donor role within them. We need to consider evaluation in this context more carefully than we often do.

Also, you can't be a development practitioner without being confronted with the importance of behaviour change by individuals, collectivities, economic actors and governments. There are some technical fixes, such as vaccination – though even in that case people have to be prepared to take the technical fix.

But most objectives require more or less significant change of behaviour, from improving your diet and taking more exercise to stemming losses from corruption, fraud and

¹ Center for Global Development (CGD), 2006. *When will we ever learn? Improving lives through impact evaluation*. Washington: CGD.

incompetence. So I'm interested in what encourages behaviour change. In 2009, I wrote a paper, 'Do Indicators Change Behaviour?' in which I assessed the MDGs from this perspective – with hindsight, maybe a shade too critically.

Can evaluation – and particularly impact evaluation – change the behaviour of policymakers? We all know about the famous evaluation of PROGRESA, which led to it surviving a change of government in Mexico and its rebranding. But the fact that we still quote this example may indicate that it is quite exceptional.

Addressing the issue becomes increasingly important as the volume of impact evaluations rises. 3ie has been documenting the impact of its studies for some time. Anecdotally, it is clear that the picture is mixed. 3ie can point to some direct consequences – such as the rolling out of pre-school programmes in Mozambique, or a change of policy for groundwater access for farmers in West Bengal. But it's clear that we cannot yet make confident statements about the overall policy impact of 3ie's portfolio, or of different products within it. I suspect that those of you from other agencies and other parts of the evaluation community face similar problems.

So let me consider with you, based on my own experience, a few ways in which we may hope to enhance the likelihood that evaluations – and particularly impact evaluations and systematic reviews -- will feed into policy.

1. Quality matters

One poorly done study may be dangerously influential, but evaluation units that produce or oversee the production of methodologically poor work should not expect to be influential over time. 3ie, for one, should continue to enforce rigorous standards of work without limiting itself to one methodology. Maybe the institutions of all bilateral and multilateral agencies should submit themselves to periodic independent or peer reviews of their own quality standards, which would be placed in the public domain.

2. Country ownership matters

Policymakers are much more likely to pay attention to evaluations in whose commissioning they have played a role (though this role must not weaken the commitment to quality and independence of judgement of the evaluation). That's not to deny that there is a role for evaluations commissioned by independent evaluation and scrutiny units without the agreement (or possibly against the wishes) of senior policymakers or other stakeholders.

I recall the role of the IMF's Independent Evaluation Office under its first head, Montek Singh Ahluwalia, who put in place a small number of evaluations of a few key and contested topics, to good effect. The UK's Independent Commission for Aid Impact is another independent institution that sets its own priority topics for review. However, I think that the point remains relevant.

One obvious example – stemming from a point I made earlier – is that, in the field of activities involving international aid, evaluations are far too often commissioned by donors and not enough by the governments responsible for much of the activities, or at least the

context in which the activities operate. We need to turn the telescope round from getting development agencies to improve (good though that would be!) to getting outcomes to improve. That means that local stakeholders need to have a much larger role, for example of the kind that we in 3ie see so clearly in members like China, Mexico or South Africa.

To its credit, 3ie is increasingly assisting this greater engagement through country policy windows and through its other services to members. Its new strategy, recently approved by its board, will further enhance its flexibility in delivering services relevant to countries at different levels of development that wish to move towards assessing the outcomes attributable to their programmes, whether or not supported by donors.

I mentioned the example of China. While China has, I believe, commissioned rather few formal impact evaluations as such (though 3ie has played a very positive role in encouraging more rigorous methods in China), China does have a longstanding policy of piloting new approaches before attempting to roll them out across such a large and diverse country. I happened to have the opportunity recently of helping supervise a paper produced by a senior Chinese researcher at the Development Research Center of the State Council. His topic was to understand how China had managed to stem losses in its huge rural cooperative credit system.

It's a fascinating story, including recognition of the problem in a go-ahead province, a personal visit to the province by no less a person than the prime minister, the commissioning of further experimental work ahead of a return prime ministerial visit, and the eventual rolling out of an entirely new system that radically changed the incentives of managers at each level. What's more, the rollout gave other provinces a choice of options that suited the different levels of development across China, while maintaining the essential disciplines of the new system. The point I want to stress here is the ability of China in this way to create real ownership of such reforms, and how this led to effective implementation.

In the same vein, just this morning, it was great to hear the World Bank praise the value of having the government as a co-PI in studies of HIV in Malawi, and how this led back to quick policy action.

When I was chair of the OECD Development Assistance Committee, I was much involved in the Paris Declaration on Aid Effectiveness. It's not for nothing that the first 'Paris Principle' is indeed ownership. Donors have to do more than pay lip service to this concept if they wish to see sustainable progress in achieving results from evaluations.

I would like to see all donor agencies set publicly a percentage of their evaluations that will be co-commissioned with the relevant Southern authorities (and ideally led by the latter). I predict that such evaluations will be much more influential in the countries concerned, and hopefully also more honest about what both donors, as well as local authorities, could do better.

3. Ownership by communities of practice matters

It's not just a matter of a better balance between donors and Southern governments. There's also the issue of buy-in from the technical experts who advise policymakers: Jim Kim's 'Communities of Practice' and their counterparts in governments and development agencies.

One feature of the globalising world is the development of increasingly effective and coordinated international communities of practice. The OECD is one good example of this tendency. The value of this approach is shown by the number of non-members who these days participate in its various committees. Local context remains vitally important, but 'what works' is often something that can guide practice in many different environments. Similarly, the donor community has invested heavily in 'special purpose funds' from longstanding ones, such as UNICEF, UNFPA, IFAD and the Global Environment Facility, to this century's suite of health sector vertical funds, the Green Climate Fund and the Global Partnership for Education.

Looking at 3ie product lines, one can see that evidence gap maps, thematic windows and systematic reviews in particular – and also the Development Priorities Window – all have particular relevance to such communities of practice. Indeed, 3ie already has links to some of these communities, for example the Water Supply and Sanitation Collaborative Council.

I think that, as the 3ie board discussed yesterday, 3ie needs to work more closely with these communities of practice, as well of course with donors prepared to fund work sought and valued by the communities of practice themselves. This should not be impossible, as most funding agencies have staff forming part of many of these communities of practice. One problem is that the communities of practice may themselves fail to give sufficient attention to the potential value to themselves of, for example, systematic reviews, and 3ie probably needs to do a better job of promoting such approaches.

As an example, donors have spent a huge amount on attempting to develop improved cookstoves to tackle indoor air pollution. An early 3ie study of one initiative in Ghana demonstrated that not all such initiatives are successful.² One study cannot be generalised. I therefore looked online for a related systematic review. There are not many. I did find a couple that look at some of the factors that influence choice of fuel and technology. One done in 2012 looks at take-up.³

That review concluded: 'Adoption studies of improved cookstoves or clean energy are scarce, scattered, and of differential quality, even though global distribution programs are quickly expanding. Future research should examine an expanded set of contextual variables to improve implementation of stove programs that can realize the "win-win-win" of health, local environmental quality, and climate associated with these technologies.'

² Burwen, J and Levine, DI, 2012. *A rapid assessment randomised-controlled trial of improved cookstoves in rural Ghana, Impact Evaluation Report 2*. New Delhi: International Initiative for Impact Evaluation. Available at: <http://www.3ieimpact.org/en/evidence/impact-evaluations/details/99/> [Accessed 16 May 2017].

³ Lewis, JJ and Pattanayak, SK, 2012. Who adopts improved fuels and cookstoves? A systematic review. *Environmental Health Perspectives*, 120(5), 637–645.

That's not exactly a ringing endorsement. Yet, so far as I am aware, donors still seem minded to conclude that the way forward is to 'try harder' when other evidence suggests that improved cookstoves, even when adopted, seldom deliver the health benefits deduced from experiments under laboratory conditions. I think that there is a strong case for a new systematic review on the health impact of cookstove programmes in practice, preceded of course by an evidence gap map to assess what evidence exists for the various issues in the theory of change. I would encourage the donors to the Cookstoves Alliance to commission such an evidence gap map as soon as possible. I am sure that Sie would be happy to oblige.

The effort to work more closely with international communities of practice and interested donors is surely worthwhile. A systematic review commissioned by one donor may or may not get serious attention from the international community of practice: one co-commissioned by the community of practice itself is much more likely to be fed back into the community concerned.

4. Feedback matters

Developing the theme of the involvement of relevant policymakers, I would also argue for better designed feedback loops. Here are two specific examples from my experience in the UK's Overseas Development Administration (ODA) in the 1980s and 1990s (at a time when aid was not particularly popular with the government of the day).

One is to get findings to people across the institution. Given that we all move around in our careers, it's worth spreading information around that may be used down the line. In those pre-digital days, it was a paper-based system, whereby staff were issued with a special binder. Each evaluation had a one-page evaluation summary that gave the highlights of every evaluation carried out by the ODA – probably in those days a dozen or so a year.

I was nerd enough to read all these – it only took a matter of minutes. That's how I first learned that water supply projects were unlikely to deliver any health benefits without some complementary investment in health information. A year or so after picking up this piece of knowledge, I was sitting in the World Bank Board as UK Alternate Executive Director looking at an engineering-based water supply project in West Africa with absolutely no health inputs at all. So at least I managed to get one point over to the mighty Bank!

It may ironically be harder to pull off such user-friendly systems in a digital age, where everyone is exposed to vast quantities of relevant and irrelevant information. But I think that it is not impossible to get out key lessons efficiently – particularly from evaluations that are more than one-off, which may of course be weak on external validity.

The second is to have the same people who approve major new programmes also look properly at the results of the evaluations that their agency at least has commissioned. In the ODA, this was done by having all the evaluations submitted to the same senior committee that advised the minister on all new commitments above the delegated authority level. If nothing else, this provided a clear feedback loop to those with authority in the system. Again, it was easier with fewer evaluations. But there should be ways of getting the results of evaluations by some means or other to the senior levels of an agency, whether a development agency or that of a government.

I also like the approach of some donors (I give the Children's Investment Fund Foundation as an example) who require all expenditure proposals to begin with a brief survey of the evaluation evidence for the intervention in question. I think that this is a piece of good practice that could be readily rolled out across agencies.

So I would like to see all development agencies (and indeed this could go much wider to government agencies in general) to require consideration of existing knowledge from evaluations before taking decisions on major new activities; to ensure that those taking major financing decisions also have a formal role in studying new evaluation findings; and that very short summaries of all evaluations, and particularly of syntheses of evaluation evidence, are widely distributed within their agencies.

5. Results matter

A final reflection on the donor community. With increasing scrutiny of the effectiveness of development spending (something that we should welcome), donors have often responded by stronger assertions about the results which this spending is achieving. These claims can often be misleading: donors taking credit for outputs (say classrooms built) even when outcomes (say better literacy of school-leavers) are not achieved; or claiming credit for outcomes even when the intervention itself had been ineffective, and the outcome is the result of progress unrelated to the project (as anyone who has listened to Howard White on the World Bank's Bangladesh Integrated Nutrition Project will readily recall). Worse, too narrow a focus on the measurable may inhibit risk-taking or the adoption of more effective but less readily quantifiable approaches.

One problem is the lack of common understanding between the impact evaluation community and those who specialise in management by (or for) results. The two should be natural allies, as both are or should be concerned at establishing whether any particular theory of change is effective in practice. Attribution of outcomes to interventions cannot be securely demonstrated without a counterfactual, so anyone interested in results management should also be an advocate of rigorous evaluation of impact.

I would like the two communities to address together ways of presenting the synergies between the insights of each in a more compelling way to policymakers.

Conclusion

As I step down from the position of chair of the 3ie board shortly, I would like to end by saying a word about the direction that 3ie is taking under its new strategy.

This strategy maintains 3ie's commitment to improving lives through evidence derived from rigorous but mixed methods, appropriate to context, and especially in areas where there are gaps in knowledge and a lack of evidence in particular countries and regions among low- and middle-income countries. I would like to stress here that 3ie is not 'an RCT factory'. The strategy has a stronger focus on take-up of evidence and includes support for a better enabling environment to generate and use high-quality evidence, including through services to the wider community such as its repository of impact evaluations, now containing almost 5,000 searchable studies.

One thing that makes 3ie special is its member-based structure. In the last 5 years, membership has grown from the low 20s to around 50, with Southern governments and institutions tripling in number from 10 to 30. It is very evident from the meetings this week, and from the many informal discussions outside the meetings, what a significant community of practice this is becoming, both among Southern members and among all members and other agencies as you have been hearing today.

In response to the growth and diversity of membership, 3ie is building in more flexible ways of cooperating with its Southern members, for example through supporting formative evaluations in countries where these are a valuable precursor to impact evaluations as such, and being ready to work with partners to help put in place systems that are needed if impact evaluations are to be successfully commissioned down the line. Consistently with what I have been saying, 3ie emphasises host country ownership, and is also experimenting with support programmes at sub-regional and sub-national level. We are also discussing how we can broaden our relevance to Northern governments and non-official institutions beyond the core group of committed organisations.

Equally, the strategy encourages the development of closer ties with international communities of practice, particularly for its synthesis products, including its innovative evidence gap maps.

3ie was created out of the concern that the world was short of evidence about the success or otherwise of major programmes. Along with other agencies, 3ie has made a lot of progress in generating and synthesising such evidence, including more recently under our Development Priorities Window in areas where rigorous evidence is particularly thin. There is still much more to do in generating evidence, but I fully support the greater focus now being given to the use of this growing body of evidence, and I see 3ie – with its longstanding commitment to linking evidence to policy – as one of the key contributors to moving this forward.

Finally, I would like to express my thanks to my fellow commissioners, who bring enormous experience, engagement and enthusiasm to guiding and promoting 3ie; to Dr Jimenez, his senior management team, and 3ie staff in its offices in Delhi, London and here in Washington for their commitment and creativity, and a particular thanks to all involved in arranging the meetings here this week; to all 3ie members, who have shown their common dedication to the generation and use of evidence; and finally to Ruth Levine, who replaces me as chair in June. Ruth's passion for usable evidence was absolutely key to 3ie's establishment, and she brings to the position of chair huge credibility among all who care about evidence-based policy.

Thank you!