Did it make a difference? Trying to unravel the mystery of impact and the contribution made by a public initiative

A paper presented at the 2009 Australasian Evaluation Society Conference
Evidence *and* Evaluation
31 August – 4 September 2009
Canberra, Australia

Author: Julie Hind, Evolving Ways

Did it make a difference? This is a question often asked by policy makers and funding bodies. It is a valid question. Most of us want public funds to make a difference. Not surprisingly, policy makers and funding bodies are increasingly looking for hard evidence that a particular intervention works – and why.

We would like to share with you our recent experiences in exploring the notion of the counterfactual in order to measure the actual difference an intervention has made as opposed to the overall change that might have occurred.

The Department of Primary Industries (DPI), Victoria commissioned us to conduct an impact evaluation of an innovation initiative two years after the completion of a particular three-year funded project. The project, which was research-based, developed a traceability system for the southern rock lobster industry. It was designed to assist in value chain improvements. The project had been something of a flagship initiative for DPI with many early performance stories published widely. Reviews and other data stemming from the implementation stage highlighted many successes: traceability was proven to be possible; a system had been established; deliverables had been met; and process outcomes had been achieved. As a consequence, there was a strong belief amongst those who had been part of the implementation that ongoing and expanded commercial success would surely have ensued.

The evaluation was intended to assess the impact of the project, including the degree to which subsequent commercialisation had been achieved. A narrative assessment of achievement against the intended outcomes hierarchy, a cost-benefit analysis, and a narrative about the lessons, were the agreed tasks.

The evaluation was not large in terms of budget and time was limited, with the commissioning agent requiring the findings in a short time period. This meant that it was to be a brief evaluation, in terms of time and methods, and limited in terms of the number of stakeholders who would participate.

Traditionally, within this program area, impact would be measured in the volume of subsequent commercial trade and through describing the extent of change through performance stories. Initial data, gathered during the scoping stage of this evaluation, identified that the subsequent commercial outcomes had not been realised to the degree anticipated. Commercial and other constraints were, largely, the contributing factors. This slow uptake of the opportunities came as a surprise to those with whom we were working and put in question the agreed evaluation focus. We felt that, given the situation, continuing with an approach that looked, largely, at performance against the outcomes hierarchy and identifying the volume of subsequent commercial sales would not deliver much in the way of value-add for the client.

Reflecting on how we might adapt the initial methodology we were guided by the principles of Patton's utilization-focused evaluation.[1] Building on the existing working relationship we had with the commissioning agent, we worked with them to review what would be of most use to them. In light of the unexpected delays in the commercialisation, the agency was interested to better understand the contribution it had made and if the program had made any difference. Had it added any longer-term value in any way?

Given that utilization-focused evaluation does not prescribe models, methods or theories, we turned to the literature to search for various value-add concepts and the different ways in which impact is evaluated.

Impact can be illustrated using the notion of the counterfactual as in the following diagrams. Broadly speaking, the impact or effect is the difference between the actual outcomes and the counterfactual: that is, what would have occurred without the policy intervention.

Diagram 1: Impact and counterfactual[2] (where the counterfactual assumes a declining situation)

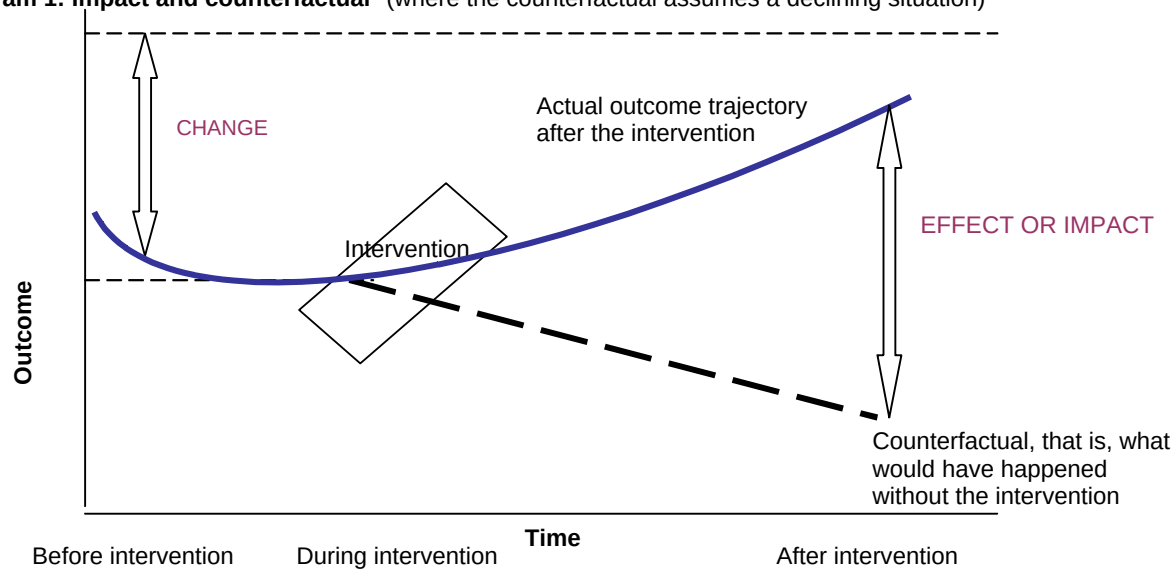
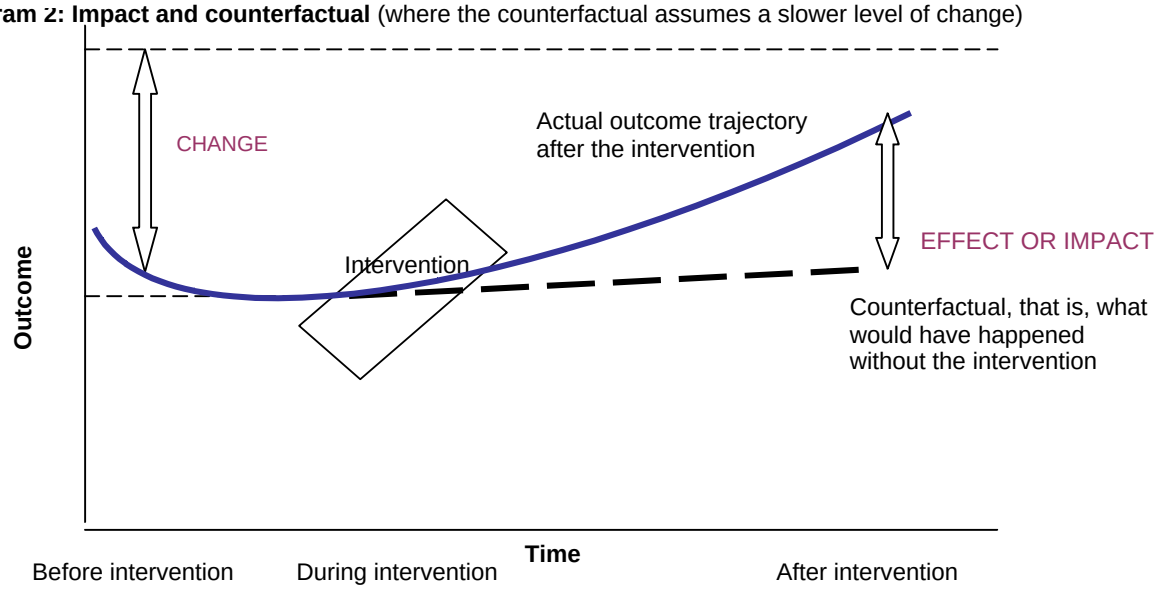


Diagram 2: Impact and counterfactual (where the counterfactual assumes a slower level of change)



[1] Patton, M. Q., *Qualitative Research and Education Methods (3rd Ed),* 2002
[2] Adapted from Lipsey, M.W., *Outcome and Impact Assessment,* The Evaluators' Institute, Washington DC, 2006.

We are aware of the debate in recent years as to what is contemporary good practice in relation to impact evaluation. 'Best practice' impact evaluations, according to the World Bank, must estimate the counterfactual and assess the difference using quantitative measures, preferably randomised control groups because these are considered more robust.[3] [4]

Whilst acknowledging other methodologies, including qualitative techniques, these are seen as enhancing "…the findings of an impact evaluation…[but that]…without a comparison group, it is impossible to determine the counterfactual and thus causality…"[5] Generally, this view is also held by the Asian Development Bank.[6]

On the other hand, DFID (the UK Department for International Development), embraces qualitative methods more readily in their own right whilst advocating a combination of quantitative and qualitative for impact studies. In a recent paper it was suggested that "…contextual studies should not be used simply to confirm or 'window dress' the findings of non-contextual surveys. Instead they should play a more rigorous role of observing and evaluating impacts, even replacing, when appropriate…"[7][traditional quantitative methods].

Likewise, AusAID acknowledges the valid place in impact evaluation for both quantitative and qualitative methods, separately or combined.[8]

The counterfactual is used to help establish causality. Scott Bayley[9], drawing on John Stuart Mill, reminds us that for causal analysis, three criteria must be satisfied: i) that there is a correlation between A and B; ii) that there is a temporal order where A occurs before B; and iii) all non-program explanations must be eliminated.

Mayne,[10] acknowledging that it is not always possible to show direct causality, shows us that uncertainty around non-program explanations can be reduced through contribution analysis by identifying the plausible association between effort and effect.

Favouring the view that impact can be assessed using a range of methods, our determination was to find a more rigorous, practical method to address particular evaluation context with which we were faced: the impact evaluation had not been built into the program design; no actual counterfactual had been developed; no controlled sample had been established and perhaps was not even relevant; the sample population was small; no longitudinal data had been gathered; the project involved a relatively small number of people; and we were constrained by time and resources. Nonetheless, our client wanted an answer to what contribution the project had made to the wider effort.

[3] Baker, J., *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners,* World Bank, Washington, 2000.
[4] World Bank, *Impact Evaluation – Overview,* www.worldbank.org
[5] Baker, J., op. cit. pg. 8.
[6] Asian Development Bank, *Impact Evaluation: Methodological and Operational Issues,* Sep. 2006
[7] Garbarino, S., and Holland, J., *Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results: Issues Paper,* Governance and Social Development Resource Centre, DFID, Mar 2009, pg. 1.
[8] Office of Development Effectiveness, *Impact evaluation,* AusAID, 2006.
[9] Bayley, S., *Evidence Requirements for Evaluating Program Impacts,* Workshop to Department of Primary Industries, 2006, slide 12.
[10] Mayne, J., "Addressing attribution through contribution analysis using performance measures sensibly", *The Canadian Journal of Program Evaluation,* 2001, Vol. 16, No. 1, pp. 1-24.

This context is not an unusual one for us. As evaluators we promote, as other evaluators do, the building of impact evaluation into the initial program design. However, the reality is that this is often not done and in some instances, even if it were, some programs are not large enough to warrant a highly sophisticated, longitudinal, controlled evaluation.

Given the research-based nature of the innovation program we were evaluating, we were interested to see how impact is conceptualised in similar fields. The contribution to change made by public intervention in the research and innovation field is described by Georghiou and his colleagues[11] as additionality, and they describe it as having three manifestations:

a. Input additionality, which seeks to assess whether the input resources are additional to what would be invested by the collaborator and not merely replace resources;

b. Output additionality, which seeks to assess the proportion of outputs that would not have been achieved without public sector support. This category pushes beyond immediate outputs to outcomes additionality and includes unintended effects and spill-overs; and

c. Behavioural additionality, which seeks to assess scale, scope and acceleration, as well as long-term changes in behaviour at the strategic level or in competencies gained.

For our evaluation, this framework seemed to provide a useful starting point for three reasons. Firstly, the nature of innovation projects in public-private collaborations (as was the case with the project we were evaluating) is that innovation does not just exist for the time of the project. It can begin before the contracted work and/or continue after the contracted work has finished, and the innovations are generally integrated with other activities that are privately funded. Georghiou emphasised that the question therefore is: What did the publicly supported contract contribute to the wider effort? This question sat very nicely with what we were seeking to discover. Clearly, the project that we were evaluating was only one part of a greater whole.

Secondly, additionality is measured through comparison with the counterfactual, and often with a combination of methods. Thirdly, considerable work has been done internationally by an OECD Working Group[12] in recent years to explore the concept of, and how to measure, behavioural additionality. Therefore, measuring impact through additionality has some solid grounding.

We used the three manifestations, which we viewed as a possible typology, and the broader notion of the counterfactual to provide additional robustness to our evaluation. Firstly, specific interview questions, aligned to each of the additionality categories, were developed. These were used to gather more precise information about the additional benefit that ensued from the project.

Secondly, not having access to a strictly defined counterfactual we, nonetheless, felt it important to have stakeholders describe, from their perspective, the expected situation had the intervention not occurred. In this way we were able to build up a counterfactual we felt was practical and sufficient for the particular context in which we were asked to work. This enabled us to explore, through interviews designed using the typology, what was different as a result of the project. Although a baseline interview prior to the intervention was not possible in this instance, we felt that this reflexive comparison was, nonetheless, valuable.

[11] Georghiou, L., *Impact and Additionality of Innovation Policy,* paper presented to Six Countries Programme on Innovation Spring Conference, 2002

[12] OECD, *Government R&D Funding and Company Behavior: Measuring Behavioural Additionality,* 2006.

A further critical component was the exploration, for each of the categories, the non-project explanations for changes and outcomes.

A key measure of success of evaluation for us is the degree to which a client finds an evaluation useful. In this instance, the client found it extremely useful. The traditional use of using volume of commercial sales as a measure of success of intervention for innovation projects does not seem to be a reliable measure of impact. This expectation of a causal link between a longer-term business initiative and a single sponsored project is described in the literature as project fallacy.[13] The additionality typology has provided a workable and practical alternative to measuring impact for this particular project. In addition, the typology has provided the program area with a conceptual framework in which to consider its work in general and how it might be used in future at the front end of a project. Should this be done in the future, it could help strengthen the design of projects.

Building on this initial successful experience, we are currently working with the client to evaluate one of its program areas using the additionality typology as one of the tools. With an interest in adding further value ourselves, we have used this opportunity to explore an expanded typology. Borrowing from the work of Carr et al[14], the International Finance Corporation[15]and the work of the OECD Working Group[16] we have expanded our working typology. This expanded typology, which we have called *The Unique Benefit*, consists of eight categories of additionality:

- Inputs;
- Outcomes;
- Behavioural;
- Quality – where the quality of the outcomes may be different because of the public sector intervention;
- Non-financial risk mitigation – where the risk perception is minimised because of public sector reputation, networks and credibility;
- Knowledge and innovation – where the public sector adds global, technical or industry knowledge and innovation not readily available elsewhere;
- Standard setting – where the public sector experience in protocols and standards adds value not readily available elsewhere; and
- Policy – where policy support and advice is provided that is not readily available elsewhere.

This particular evaluation is currently in progress so the usefulness of this expanded typology is yet to be fully tested. However, initial reactions from the evaluation team suggest that *The Unique Benefit* typology is assisting in gaining a greater breadth and depth of understanding of the impact of the program.

Georghiou and Clarysse[17] note that considerable work needs to be done to find reliable methodological approaches to measuring additionality. Whilst our effort to contribute to this body of work is small, and it is still a work in progress, we do feel that *The Unique Benefit* has promise.

[13] Georghiou, L., and and Clarysse, B., "Introduction and Synthesis", in OECD, *Government R&D Funding and Company Behavior: Measuring Behavioural Additionality*, 2006.
[14] Carr, S., Dancer, S., Russell, G., Tyler, P., Robson, B., Lawless, P., and Haigh, N, *Additionality Guide: A Standard Approach to Assessing the Additional Impact of Projects,* (Second edition), English Partnerships, Sep. 2004.
[15] International Finance Corporation, *IFC's Role and Additionality: A Primer,* The World Bank Group, May 2009.
[16] OECD, 2006, op. cit.
[17] Georghiou L., and Clarysse, B., op. cit.

We feel that *The Unique Benefit* can be one beneficial method in the armoury of impact evaluation and that it can contribute to unravelling the mystery of impact and contribution of public initiatives.

References

Asian Development Bank, *Impact Evaluation: Methodological and Operational Issues,* Sep. 2006.

Baker, J., *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners,* World Bank, Washington, 2000.

Bayley, S., *Evidence Requirements for Evaluating Program Impacts,* Workshop to Department of Primary Industries, 2006.

Carr, S., Dancer, S., Russell, G., Tyler, P., Robson, B., Lawless, P., and Haigh, N, *Additionality Guide: A Standard Approach to Assessing the Additional Impact of Projects,* (Second edition), English Partnerships, Sep. 2004.

Garbarino, S., and Holland, J., *Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results: Issues Paper,* Governance and Social Development Resource Centre, DFID, Mar 2009.

Georghiou, L., *Impact and Additionality of Innovation Policy,* paper presented to Six Countries Programme on Innovation Spring Conference, 2002

Georghiou, L., and and Clarysse, B., "Introduction and Synthesis", in OECD *Government R&D Funding and Company Behavior: Measuring Behavioural Additionality,* 2006.

International Finance Corporation, *IFC's Role and Additionality: A Primer,* The World Bank Group, May 2009.

Lipsey, M.W., *Outcome and Impact Assessment,* The Evaluators' Institute, Washington DC, 2006.

Mayne, J., "Addressing attribution through contribution analysis using performance measures sensibly", *The Canadian Journal of Program Evaluation,* 2001, Vol. 16, No. 1, pp. 1-24.

OECD, *Government R&D Funding and Company Behavior: Measuring Behavioural Additionality,* 2006.

Office of Development Effectiveness, *Impact evaluation,* AusAID, 2006.

Patton, M. Q., *Qualitative Research and Education Methods (3rd Ed),* Sage, 2002.

World Bank, *Impact Evaluation – Overview,* www.worldbank.org