

# Additionality: a useful way to construct the counterfactual qualitatively?

## Abstract

One concern about impact evaluation is whether a program has made a difference (Owen 2006). Essentially, in trying to understand this, an evaluation seeks to establish causality—the causal link between the social program and the outcomes (Mohr 1999; Rossi & Freeman 1993). However, attributing causality to social programs is difficult because of the inherent complexity and the many and varied factors at play (House 2001; Mayne 2001; Pawson & Tilley 1998; White 2010).

Evaluators can choose from a number of theories and methods to help address causality (Davidson 2000). Measuring the counterfactual—the difference between actual outcomes and what would have occurred without the intervention—has been at the heart of traditional impact evaluations. While this has traditionally been measured using experimental and quasi-experimental methods, a counterfactual does not always need a comparison group (White 2010) and can be constructed qualitatively (Cummings 2006).

With these in mind, this article explores the usefulness of the concept of *additionality*, a mixed-methods framework developed by Buisseret et al. (cited in Georghiou 2002) as a means of evaluative comparison of the counterfactual.

## Introduction

Over recent decades, there has been a move away from an emphasis on procedural matters within government agencies towards a focus on achievements (Mayne 2001). It is important for governments to demonstrate that they have used public funds responsibly to achieve desired results, also known as ‘outcomes’ (Patton 2002; Schweigert 2006; White 2010).

This changing emphasis has ‘... raised the bar for what evaluations should produce’ (Schweigert 2006, p. 416) because it brings with it a demand for evaluations to do more than judge whether a program has been implemented well. Rather, evaluations are increasingly focused on: whether outcomes were achieved;

## Julie Hind



*Julie Hind is Director of Evolving Ways, Albury-Wodonga. Email: <julie.hind@evolvingways.com.au>*

*This article was written towards completion of a Master of Assessment and Evaluation degree at the University of Melbourne.*

whether there were any unintended outcomes (both positive and negative); and whether the program made a difference. Each of these questions falls within the definition of impact evaluation as defined by Owen (2006). However, questions about whether a program has made a difference have a particular nuance. They seek answers about attribution, that is, the causal link between program inputs, activities and outcomes (Mayne 2001; Mohr 1999; Rossi & Freeman 1993).

At the same time, there is widespread agreement that attributing causality to social programs is difficult (House 2001; Mayne 2001; Pawson & Tilley 1998; White 2010). Factors such as other government actions, social trends and the economic situation can have an effect on outcomes (Mayne 2001). Likewise, the complexity inherent in social programs has an effect (Pawson & Tilley 1998). As a consequence, it is hard to be certain of all the social events involved (House 2001), making it difficult to be definitive about program effects. However, despite the many challenges, it is essential for evaluators to find appropriate ways to determine causality (Davidson 2000) because an important contribution of evaluation to the improvement of the social condition is to assess project effects (Lipsey 2001). If the question of causality is left unanswered, ‘... little can be said about the worth of the program, neither can advice be provided about future directions’ (Mayne 2001).

### Ways to address causality

Evaluators can choose from a number of theories and methods to address causality (Davidson 2000). Traditionally, impact evaluation has been conducted using experimental and quasi-experimental studies where the outcomes of intervention groups are compared with control groups to determine the difference made by an intervention.

This difference between actual outcomes and what would have occurred without an intervention is referred to as the ‘counterfactual’. The counterfactual is at the heart of traditional impact evaluations. For some evaluators, well designed and executed evaluations that use experimental and quasi-experimental methods are considered to be the best way to judge the effects of any program (Lipsey 2001). Others consider that they are the best methods for large population-based programs (White 2001).

However, over time, other methods have been developed in order to address some of the perceived problems with experimental and quasi-experimental methods. For example, ‘realist evaluation’ was developed to help address a perceived theoretical issue, namely, that traditional approaches fail to explore *why* a program works or fails (Mark & Henry 1998; Pawson & Tilley 1998). Realist evaluation therefore tests and refines hypotheses about mechanisms, contexts and outcomes in an attempt to understand what works for whom and in what circumstances. Understanding mechanisms and what makes a program work is also the

focus of theory-based evaluations (Davidson 2000). In addition, theory-based methods have responded to perceived methodological challenges of experimental methods by recognising that the various assumptions underlying a program need not be controlled but just acknowledged and explored (Rogers et al. 2000). A further example is that of ‘real-world evaluation’, which has emerged in response to time, budget and data constraints that can make traditional experimental methods impractical (Bamberger 2006). This approach promotes a more flexible design through the selection of the ‘... strongest possible design within the particular set of budget, time and data constraints’ (Bamberger 2006, p. 2).

The particular concept that is the subject of this article is that of *additionality*, developed by Buisserat et al. (cited in Georghiou 2002) in response to difficulties with using experimental methods to evaluate research and development initiatives. The term ‘additionality’ involves a mixed-methods approach of evaluative comparison with the counterfactual (Georghiou 2002; Georghiou & Clarysse 2006).

I have applied the additionality concept in several evaluations over the past two years with some degree of success. This article puts the case for the use of additionality alongside other methods such as theory-based and realist evaluation, as an effective lens for seeking out the unique benefits of a program that might otherwise not be uncovered readily. In putting the case, I begin by outlining the importance of causality in program evaluation. The article then describes the counterfactual and how this can be applied qualitatively as well as quantitatively. Additionality is then outlined further, with attention given to how and why it can be applied. The final section of the article considers in which particular evaluation situations additionality can be a useful concept, as well as how it can help to increase the influence of an evaluation (Kirkhart 2000).

### Causality—its relevance to program evaluation

So why is program evaluation concerned with issues of causality? While not always concerned with questions of causality—for example, in proactive forms and some aspects of interactive and monitoring forms of evaluation (Owen 2006)—much of the work of program evaluation has causality at its core. For instance, wherever program evaluation seeks to understand how a program works in theory, rather than simply how it was or should be implemented, questions of causality arise.

There is also a growing demand to account for public funds by demonstrating that programs make a difference (Mayne 2001; Schweigert 2006), with ‘... government and the sceptical public ... less interested in reports on the effectiveness of program implementation, and the tallying of outputs, and instead [demanding] to know about program

impact' (Kotvojs & Shrimpton 2007, p. 27). Therefore, being able to infer causality is important.

*Causality* is defined in the *Macquarie Dictionary* (2009) as 'the relation of cause and effect' where *cause* is further defined as: 'that which produces an effect; the thing, person, etc., from which something results'. While easily understood and recognised in our daily lives, a precise definition of causality has been debated by philosophers and scientists for centuries (Scriven 1991; Shadish, Cook & Campbell 2002; Schweigert 2006).

Even so, Hume is considered as the single most influential thinker about causality (Mackie 2002). It is his concept, known as the 'Hume's theory of causation', that the evaluation field has inherited (House 2001). This theory is based on the notion of regularity of succession, that is, where one event takes place before another repeatedly, there is reason to believe these events will occur together again (House 2001). This successionist view of causation attributes the cause externally to the object, that is, the effect occurs because of some force external to the object (Pawson & Tilley 1998). Further, it is external because the relationship between cause and effect cannot be perceived directly; rather, it must be inferred (Schweigert 2006).

Over time, causation, as it applies to social programs, has been recognised as more complex than Hume's theory suggests (House 2001) and has given rise to generative concepts of causation. These highlight the importance of explanation and the need to understand the mechanisms by which causes influence effects (House 2001; Mackie 2002; Mark & Henry 1998; Pawson & Tilley 1998; Shadish, Cook & Campbell 2002). As well as acknowledging the part played by external forces, generative concepts emphasise the importance of internal attributes of cause, such as the willingness of the participants and the context in which the program is implemented (Pawson & Tilley 1998).

While theorists debate which of the concepts is the more relevant, robust and rigorous, causality is not just an academic issue (Davidson 2000). It is a real and tangible issue for the practitioner when asked to judge the merit and worth of a program based on its impact. However, the preferred method for assessing impact, and therefore inferring causality, has likewise been subject to debate. Traditional methods have been experimental and quasi-experimental designs, which are built on Hume's successionist view of causation (House 2001). They focus on '... manipulation, control and observation' (Pawson & Tilley 1998, p. 33) so that inferences can be made about the link between cause and effect. Advocates for these tightly controlled quantitative methods believe that such processes offer greater internal validity than other methods because they increase the confidence that the program is indeed the cause of the outcomes (Campbell & Russo 1999).

On the other hand, approaches, such as theory-based and realist evaluation, that are based on the generative view of causation, attempt to understand how, and why, programs work (Davidson 2000;

Pawson & Tilley 1998). These approaches do not control for variables but rather explore, in context, the interactions between people and events (House 2001). To help determine attribution, these methods delineate the many components, assumptions and underlying mechanisms of programs, and then test them with a wide range of stakeholders and other sources of evidence (Blamey & Mackenzie 2007). These approaches are more flexible in the methods chosen, with practitioners choosing qualitative, quantitative or mixed methods. These approaches, based on the generative view of causality, do not set out to prove attribution definitively but rather seek to make judgements about the plausibility of such claims (Julnes & Mark 1998).

Nevertheless, there is a large body of opinion that states that experimental methods, or random-controlled trials, are the best methods for determining causality (Lipsey 2001; Shadish, Cook & Campbell 2002).

However, while randomised control trials '... can *in concept* produce stronger [causal] evidence' (Gargani 2010, p. 131), they are not easy to implement well and in practice are not always practical or possible (Bamberger, Rugh & Mabry 2006).

Whatever the point of view, it is important that evaluation practitioners have methods that can be used regularly, not occasionally (Shadish, Cook & Campbell 2002). As a result, many methodologies have been developed over the past few decades to help address the difficulties associated with experimental approaches. Realist evaluation (Mark, Henry & Julnes 1998; Pawson & Tilley 1998), theory-based evaluation (Davidson 2000; Rogers 2000; Weiss 2000;), contribution analysis (Mayne 2001), and modus operandi (Scriven, cited in Shadish, Cook & Campbell 2002) are a few of these. The long, and seemingly never-ending, search for practical approaches is important because as Lipsey (2001) notes '... there is no issue we face that is more critical to our role as evaluators than that of finding viable methods for assessing routinely the impact of everyday practical programs on the social conditions they address' (p. 326). It was in response to just such a need for a practical and viable method to assess impact regularly, that my colleagues and I turned to the concept of additionality.

### **Application of additionality to a case**

The Department of Primary Industries (DPI), Victoria, commissioned the team with whom I work to conduct an impact evaluation of an innovation initiative two years after the completion of what had been a 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. Moreover, 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. Consequently, there was a strong belief among program staff that ongoing and expanded commercial success would surely ensue.

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 learnt, 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, as well as being limited in terms of the number of stakeholders that would participate.

Traditionally, within this particular program area, impact would be measured by the volume of subsequent commercial trade and through describing the extent of overall change through performance stories. Initial data, gathered during the scoping stage of the evaluation, identified that the subsequent commercial outcomes had not been realised to the degree anticipated. Commercial and other constraints were, largely, the contributory 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. Given that this project had been something of a flagship for the agency, there was a high level of what Weiss (2000) notes as the internal politics involved. In this instance, an evaluation that measured against the outcomes hierarchy and identified the volume of subsequent commercial sales would not deliver anything of real value to the agency and the results would be unpalatable. Unless our revised methodology sought to address these process use issues, there was a risk that the evaluation would be a waste of time.

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 (Patton 2008). The agency was interested to understand better: if its contribution had made any difference; whether the program added any longer term value; the reasons behind the poor uptake; and what, if anything, could have been tackled differently.

### Why additionality and what is it?

Traditionally, additionality has been a concept used by government policymakers and administrators to analyse project viability through different perspectives as a means of complementing the essential financial viability perspective; in other words, to assess whether proposed projects will create, generate and add value (Roldán 2002). For example, it has been used by the International Finance Corporation to help gain insight into how public funds can help stimulate development funded by multilateral development banks (IFC 2008).

A review of the literature found that the concept of additionality has been used more recently to evaluate research and innovation projects following work by Buisseret et al. (cited in Georghiou 2002; Georghiou & Clarysse 2006). Essentially, its use in this field is in ‘... measuring the effectiveness of policy instruments for stimulating [research, technology, development and innovation]’ (Närfelt & Wildberger 2006). Additionality recognises that innovation projects in public–private collaborations do not just exist for the time of the project. They can begin before the contracted work and continue after the contracted work has finished. Furthermore, innovations are generally integrated with other activities that are privately funded. Hence, Georghiou (2002) emphasised that the additionality question therefore is: *What did the publicly supported contract contribute to the wider effort?* In this research and innovation context, additionality is described as having three elements:

- *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
- *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 spillovers
- *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.

### Additionality’s relationship with the counterfactual

Additionality is based on the concept of the counterfactual and stems from work on causality by Hume (cited in Mackie 2002) and Lewis (2001). Indeed, Lewis’s work established the counterfactual as the principal approach to the analysis of causation (White 2010). In essence, the counterfactual is the net effect of an intervention, that is, the changes brought about by a program over and above what might have ordinarily have occurred (Rossi, Lipsey & Freeman 2004). The counterfactual, and therefore additionality, seeks something more specific than the overall effect of a program.

Because of its use to infer causality, the counterfactual has been at the heart of experimental or random-controlled trials. However, there is growing opinion that the counterfactual can be constructed in other ways, both quantitatively and qualitatively (Carden, Bamberger & Rugh 2009; Cummings 2006; Greene et al. 2007; Mayne 2001; Mohr 1999; Scriven 2008). Indeed, Cook et al. (2009, p. 114) have recently expressed the view that random-controlled trials are ‘... neither necessary nor sufficient for a well-warranted causal inference’.

A Think Tank held in 2009 (Carden, Bamberger & Rugh 2009) outlines an extensive range of alternatives to the conventional counterfactual, dividing them into theory-based approaches, quantitatively oriented approaches and qualitatively oriented approaches. While additionality is not one of those listed in that particular list, it is, nonetheless, one such alternative way to construct the counterfactual.

### Potential methods and approaches

Where additionality has been used, the methods to construct and explore the counterfactual have not been rigid. A number of tools have been developed, for example, self-assessment checklists (Närflet & Wildberger 2006) and matrices (Roldán 2002). Many of the studies undertaken using the concept have employed case studies or surveys (Georghiou & Clarysse 2006). Our evaluation used a mixed-method data collection approach that included: group interviews with program deliverers; one-on-one interviews with program participants; a questionnaire that sought cost–benefit information; and document analysis of program material, including periodic reviews undertaken and reports compiled during implementation.

We also used mixed methods in terms of the evaluation approach. We chose not to use additionality on its own but in conjunction with a theory-based approach. We elected to do this because additionality seeks two key answers that are similar to those sought by theory-based evaluations. First, ‘behavioural additionality’ aims to measure explicitly the changes that are expected to endure beyond a project. Second, it not only seeks to understand the effects but also the means by which these are achieved (Georghiou & Clarysse 2006). We, therefore, built up the program theory with program staff to uncover: the context within which it occurred; the changes sought by the program; the underlying assumptions of why and how it might work; and the drivers behind the program and people’s participation. In this way we built up agreement concerning which theories and which links would focus this evaluation (Weiss 2000). While building up this theory, we used the concept of additionality to help frame the discussions. While not having a formalised counterfactual, we used this process to hypothesise what the counterfactual might have been. The theory and the hypothesised counterfactual then helped to guide interviews and our analysis, particularly the exploration of non-project explanations or changes.

A key part of the analysis involved working through the preliminary findings with the commissioning agent and program staff and, together, discussing the extent to which various pieces of evidence suggested additionality. This step helped to provide greater rigour in relation to the analysis of causation and, therefore, what was additional. Consequently, confidence in the causal links was greater as a result of accepting causal claims through a process of intense scrutiny by a group of knowledgeable others (Cook et al. 2009).

The experience of this evaluation led us to explore the use of the concept more widely, particularly as some of the findings did not always fit snugly in to the three expressed factors: input, output and behavioural additionality. Borrowing from the work of Carr et al. (2004), the International Finance Corporation (2009), and the work of the OECD Working Group (OECD 2006), we have expanded our working typology. We have called this *The Unique Benefit* and it consists of eight categories of additionality—the first three, as previously stated, plus:

- *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
- *policy*—where policy support and advice is provided that is not readily available elsewhere.

Using this expanded typology we have now undertaken several more evaluations of a broader range of programs within the same agency. For each, we have found the concept a useful lens through which to explore the additional benefit of publicly funded intervention. Such transferability of the concept has been confirmed elsewhere (OECD 2006). This should not be surprising given that the underlying contexts of research and innovation are similar to all social programs, that is, they are complex social systems themselves and do not exist as stand-alone activities separated in time from other complementary activities (Pawson & Tilley 1998; Weiss 2000). Therefore, the question about what, and how, do publicly supported programs contribute to the wider effort has broader relevance beyond research and innovation.

### Advantages of additionality

So why use additionality, especially when there are many other alternative methods? Indeed why additionality when it was used in conjunction with a theory-based evaluation method? The answer is that our initial decision to use it was based on the fact that the program had many similarities with the research and innovation field and, therefore, borrowing from that field had some appeal. However, in using it we have found two key advantages. The first is that the language of additionality has provided the evaluation team and those who commissioned us with a new way to view and discuss a particular program. As Patton tells us:

The evaluation language we choose and use, consciously or unconsciously, necessarily and inherently shapes perceptions, defines ‘reality’, and affects mutual understanding. (cited in Kirkhart 2000, p. 7)

The language of additionality helped to focus on the unique contribution of the agency and the program, and provided program staff with a new perspective about their work. While additionality is not the only way to explore contribution, its particular language has helped to focus people’s thinking and exploration more specifically. It has helped people grapple with the difference between the overall effect and unique benefit of the program, something that has not been a usual practice within this particular agency given its predominant use of performance stories.

To illustrate, I will cite a few examples. The first comes from an evaluation of a program that sought to improve international market development for Victorian agri-food products. The findings indicated that for some participants their input of resources was not additional in that they had used the opportunities afforded through the program to offset their regular expenditure. For others, their inputs were clearly additional because it was new expenditure that, without the program, they would not have made, and thus not ventured into international markets. In this instance, the language of additionality helped program staff make clearer assertions about the cost–benefit of the program. The second is not specific to any one evaluation but is generic to additionality. The use of spillover as a notion of program consequence is common in this particular agency. The fact that additionality actively seeks spillovers and can frame them into a specific language of additional benefit resonated well.

Because the language is easy to comprehend, many of the program staff, through discussions with the evaluation team, have come to see that additionality is a concept that could readily be applied at the program design stage. It provides an opportunity for upfront discussions about the added value government might bring to any given new program. This thinking mirrors the current work of Närfelt & Wildberger (2006) in which they are developing checklists for use in the design and implementation of policy measures.

The language of additionality, particularly its espoused use of the counterfactual, also helps to explain why we have used it in conjunction with theory-based evaluation. While theory-based approaches will have also helped uncover the added benefits, the use of additionality helped to sharpen the focus on the program’s theory, the evaluation questions and the analysis. Its explicit take-up of the counterfactual meant that there was an extra dimension to the evaluation. While some might think this is an unnecessary addition, I am reminded of Mayne’s words (2001), which state:

... we can almost always gather additional data and information that will increase our understanding about a program and its impacts, even if we cannot ‘prove’ things in an absolute sense (p. 6) ... in trying to reduce the uncertainty surrounding attribution, using as many lines of evidence as possible is a sensible, practical and credible strategy. (p. 20)

The second advantage has been to do with evaluation influence. Almost all of the programs we have evaluated using additionality have been steeped in internal politics. Many programs were subject to internal debates about their value and a considerable number of staff felt the pressure to deliver good news stories. Many also felt vulnerable to possible negative evaluation findings. In this environment, the use of additionality provided some certainty that the added benefits of the program would be uncovered. This made the evaluations feel more balanced, and staff have been more able to take on board negative findings rather than discredit the evaluations.

Apart from this more obvious process use of the evaluation, we have found that the influence of additionality is more comprehensive and more akin to the model put forward by Kirkhart (2000). Here, integrated theory of influence incorporates three dimensions: source, intention and time frame. Additionality with its readily grasped language enabled a generative process of change, both during the evaluation and through the results (Kirkhart’s ‘source’ dimension). By offering a new way of thinking about programs, there was an unintended influence that helped parts of the agency look beyond the few evaluative methods and tools they had used until then (Kirkhart’s ‘intention’ dimension). The results of these evaluations have been used more widely by the agency than initially anticipated, suggesting an end-of-cycle influence (Kirkhart’s time frame dimension).

### Concluding statements

Leviton and Gutman (2010), warn us that ‘... there are so many unjustified claims about new evaluation practices that we need to be careful to justify exactly what is new’ (p. 8). Additionality is not new. It is a concept that has been used by policymakers and government-funded financiers to assess the potential for projects for some time. In recent years it has been used successfully in the research and innovation field. Furthermore, what it seeks to uncover is not new either. Theory-based, realistic and contribution analysis approaches also seek answers to similar questions.

However, what additionality can add to the broader evaluation field is another alternative to constructing the counterfactual that is simple to grasp and easy to use. I believe that this is of enormous benefit as many of the current alternatives are complicated to use and difficult for many program staff to comprehend. Being

simple to grasp and easy to use, is not to say that additionality is devoid of rigour. Like all credible evaluation approaches and techniques, it requires us to explore evidence broadly, think deeply and justify our conclusions professionally. It is just that its easy language and concepts are readily grasped and therefore offer a useful way to communicate to those who commission evaluations as well as to various stakeholders. Additionality is flexible in that it can be applied using a range of techniques. It can also be used in conjunction with other methods. While we have used it only with theory-based and realist evaluation, it is possible that it could be used qualitatively with random-controlled trials.

Our experience suggests that additionality also adds value in situations where program politics place evaluation utilisation at risk and where there are debates about the value of programs. Its deliberate focus on the additional benefit achieved through publicly funded intervention offers program staff the hope that the evaluation will be balanced. Notwithstanding this, its focus on the additional benefit can also provide insight into whether and how much the outcomes are a result of a program.

These comments are not made with the suggestion that no other method can achieve these outcomes. However, its easy-to-grasp language and concepts, its flexibility in terms of methods used, the fact that it can be applied where budget, time and data constraints occur, coupled with its ease of use, make additionality a useful tool for inferring causality.

Finally, I am reminded of Davidson's (2005) comments that we do not need to abandon the causal issue simply because of the challenges that surround it. She offers a very useful framework for deciding if, and how, one might design an evaluation that will help make causal inferences. First, it is important to ascertain the degree to which the client wants us to be absolutely certain about the causal link. The second is about applying the two key principles of causal inference: 1) look for evidence for and against the suspected cause; and 2) look for evidence for and against other important possible causes. The third is to identify the types of available evidence that might help identify or rule out a causal link. The final one is to decide the blend of evidence that will generate the level of certainty needed in the most cost-effective way. I believe that additionality meets this framework and while it is not the only method to do so, it is, nonetheless, a useful one.

## References

- Bamberger, M 2006, *Conducting quality impact evaluations under budget, time and data constraints*, The World Bank, Washington, DC.
- Bamberger, M, Rugh, J & Mabry, L 2006, *Real world evaluation: working under budget, time, date and political constraints*, Sage, Thousand Oaks, California.
- Blamey, A & Mackenzie, M 2007, 'Theories of change and realistic evaluation: peas in a pod or apples and oranges?', *Evaluation*, vol. 13, no. 4, pp. 439–455.
- Campbell, D & Russo, MJ 1999, *Social experimentation*, Sage, Thousand Oaks, California.
- Carden, F, Bamberger, M & Rugh, J 2009, *Alternatives to the conventional counterfactual: follow-up note on Session 713 Think Tank*, American Evaluation Association annual conference, November, Orlando, viewed 26 April 2010, <[http://www.realworldevaluation.org/Additional\\_RWE\\_materials.html](http://www.realworldevaluation.org/Additional_RWE_materials.html)>.
- Carr, S, Dancer, S, Russell, G, Tyler, P, Robson, B, Lawless, P & Haigh, N 2004, *Additionality guide: a standard approach to assessing the additional impact of projects*, 2nd edn, English Partnerships, London.
- Cook, TD, Scriven, M, Coryn, CLS & Evergreen, SDH 2009, 'Contemporary thinking about causation in evaluation: a dialogue with Tom Cook and Michael Scriven', *American Journal of Evaluation*, vol. 31, no. 1, pp. 105–117.
- Cummings, R 2006, 'What if?: the counterfactual in program evaluation', *Evaluation Journal of Australasia*, vol. 6 (new series), no. 2, pp. 6–15.
- Davidson, EJ 2000, 'Ascertaining causality in theory-based evaluation', in A Petrosino et al. (eds), *Program theory in evaluation—opportunities: New directions for evaluation*, 87, Jossey-Bass, San Francisco.
- Davidson, EJ 2005, *Evaluation methodology basics: the nuts and bolts of sound evaluation*, Sage, Thousand Oaks, California.
- Garbarino, S & Holland, J 2009, *Quantitative and qualitative methods in impact evaluation and measuring results: issues paper*, Governance and Social Development Resource Centre, Department for International Development, London.
- Gargani, J 2010, 'A welcome change from debate to dialogue about causality', *American Journal of Evaluation*, vol. 31, no. 1, pp. 131–132.
- Georghiou, L 2002, 'Impact and additionality of innovation policy', paper presented to Six Countries Programme on Innovation, Spring Conference, Brussels, 28 February – 1 March.
- Georghiou, L & Clarysse, B 2006, 'Introduction and synthesis', in OECD (ed.), *Government R&D funding and company behaviour: measuring behavioural additionality*, OECD Publishing, Paris.
- Greene, JC, Lisey, MW, Schwandt, TA, Smith, NL & Tharp, RG 2007, 'Method choice: five discussant commentaries', in GJ Julnes & DJ Rog (eds), *Informing federal policies on evaluation methodology—building the evidence base for method choice in government sponsored evaluation: New directions for evaluation*, 113, Jossey-Bass, San Francisco.
- House, E 2001, 'Unfinished business: causes and values', *American Journal of Evaluation*, vol. 22, no. 3, pp. 309–315.
- International Finance Corporation 2008, *Evaluating private sector development in developing countries: assessing the additionality of multilateral development banks—IFC's experience*, IFC, The World Bank Group, Washington, DC.
- International Finance Corporation 2009, *IFC's role and additionality: a primer*, IFC, The World Bank Group, Washington, DC, viewed 5 July 2009, <[http://www.ifc.org/ifcext/about.nsf/AttachmentsByTitle/AdditionalityPrimer/\\$FILE/AdditionalityPrimer.pdf](http://www.ifc.org/ifcext/about.nsf/AttachmentsByTitle/AdditionalityPrimer/$FILE/AdditionalityPrimer.pdf)>.
- Julnes, G & Mark, MM 1998, 'Evaluation as sensemaking: knowledge construction in a realist world', in GT Henry, G Julnes & MM Mark (eds), *Realist evaluation: an emerging theory in support of practice: New directions for evaluation*, 78, Jossey-Bass, San Francisco.

- Kirkhart, KE 2000, 'Reconceptualizing evaluation use: an integrated theory of influence', in VJ Caracelli & H Preskill (eds), *The expanding scope of evaluation use: New directions for evaluation*, 88, Jossey-Bass, San Francisco.
- Kotvojs, F & Shrimpton, B 2007, 'Contribution analysis', *Evaluation Journal of Australasia*, vol. 7, no. 1, pp. 27–35.
- Leviton, LC & Gutman, MA 2010, 'Overview and rationale for the systematic screening and assessment method', in LC Leviton, LK Khan & N Dawkins (eds), *The systematic screening and assessment method: finding innovations worth evaluating: New directions for evaluation*, 125, Jossey-Bass, San Francisco.
- Lipsey, MW 2001, 'Unsolved problems and unfinished business', *American Journal of Evaluation*, vol. 22, no. 3, pp. 325–328.
- Lewis, DK 2001, *Counterfactuals*, Blackwell, Maiden, Massachusetts.
- Mackie, JL 2002, *The cement of the universe: a study of causation*, Oxford Press, Oxford.
- Mark, MM & Henry, GT 1998, 'Social programming and policy-making: a realist perspective', in GT Henry, G Julnes & MM Mark (eds), *Realist evaluation: an emerging theory in support of practice: New directions for evaluation*, 78, Jossey-Bass, San Francisco.
- Mark, MM, Henry, GT & Julnes, G 1998, 'A realist theory of evaluation practice', in GT Henry, G Julnes & MM Mark (eds), *Realist evaluation: an emerging theory in support of practice: New directions for evaluation*, 78, Jossey-Bass, San Francisco.
- Mayne, J 2001, 'Addressing attribution through contribution analysis using performance measures sensibly', *The Canadian Journal of Program Evaluation*, vol. 16, no. 1, pp. 1–24.
- Mohr, L 1999, 'The qualitative method of impact analysis', *American Journal of Evaluation*, vol. 20, no. 1, pp. 69–84.
- Närfelt, K-H & Wildberger, A 2006, 'Additionality and funding agencies: opening the black box', paper presented at the New Frontiers in Evaluation Conference, Vienna, 24–25 April.
- OECD 2006, 'Executive summary', in OECD (ed.), *Government R&D funding and company behaviour: measuring behavioural additionality*, OECD Publishing, Paris.
- Owen, J 2006, *Program evaluation: forms and approaches*, 3rd edn, Allen & Unwin, Crows Nest, New South Wales.
- Patton, MQ 2002, *Qualitative research and education methods*, 3rd edn, Sage, Thousand Oaks, California.
- Patton, MQ 2008, *Utilization-focused evaluation*, 3rd edn, Sage, Thousand Oaks, California.
- Pawson, R & Tilley, N 1998, *Realistic evaluation*, Sage, Thousand Oaks, California.
- Rogers, PJ 2000, 'Causal models in program theory evaluation', in A Petrosino et al. (eds), *Program theory in evaluation: challenges and opportunities: New directions for evaluation*, 87, Jossey-Bass, San Francisco.
- Rogers, P, Petrosino, A, Huebner, TA & Hacsí, TA 2000, 'Program theory evaluation: practice, promise, and problems', in A Petrosino et al. (eds), *Program theory in evaluation: challenges and opportunities: New directions for evaluation*, 87, Jossey-Bass, San Francisco.
- Roldán, J 2002, 'Designing a practical instrument for measuring developmental additionality: the IIC approach', paper presented at inter-agency roundtable on additionality of private sector development programs and operations supported by the international financial institutions, Inter-American Investment Corporation, Washington, DC, 23 May.
- Rossi, PH & Freeman, HE 1993, *Evaluation: a systematic approach*, 5th edn, Sage, Newbury Park, California.
- Rossi, PH, Lipsey, MW & Freeman, HE 2004, *Evaluation: a systematic approach*, 7th edn, Sage, Thousand Oaks, California.
- Schweigert, F 2006, 'The meaning of effectiveness in assessing community initiatives', *American Journal of Evaluation*, vol. 27, no. 4, pp. 416–436.
- Scriven, M 1991, *Evaluation thesaurus*, 4th edn, Sage, Newbury Park, California.
- Scriven, M 2008, 'A summative evaluation of RCT methodology: an alternative approach to causal research', *Journal of MultiDisciplinary Evaluation*, vol. 5, no. 9, pp. 11–24.
- Shadish, WR, Cook, TD & Campbell, DT 2002, *Experimental and quasi-experimental designs for generalized causal inference*, Houghton Mifflin, Boston.
- Weiss, CH 2000, 'Which links in which theories shall we evaluate?', in A Petrosino et al. (eds), *Program theory in evaluation: challenges and opportunities: New directions for evaluation*, 87, Jossey-Bass, San Francisco.
- White, H 2010, 'A contribution to current debates in impact evaluation', *Evaluation*, vol. 16, no. 2, pp. 153–164.