



A utilization-focused approach to contribution analysis

Evaluation

18(3) 364–377

© The Author(s) 2012

Reprints and permission: sagepub.

co.uk/journalsPermissions.nav

DOI: 10.1177/1356389012449523

evi.sagepub.com



Michael Quinn Patton

Independent Evaluation Consultant, USA

Abstract

Utilization-focused evaluation involves identifying and working with primary intended users to design and interpret an evaluation. This includes the process of working with primary intended users to render judgments about the extent to which the preponderance of evidence supports a meaningful and useful conclusion about degree to which an intervention has affected observed outcomes and impacts. This is the essence of contribution analysis. Two in-depth examples illustrate this process.

Keywords

contribution analysis, utilization-focused evaluation

Introduction

Contribution analysis is founded on theory-based approaches to evaluation. An explicit program theory (Funnell and Rogers, 2011) that specifies hypothesized causal linkages between the intervention and desired outcomes provides a solid foundation for analyzing both attribution and contribution pathways. One important source of hypotheses about attribution and contribution is the primary intended users of an evaluation. What is their perspective on, expectations about, and theory regarding the linkages between intervention activities and outcomes? A pragmatic, utilization-focused stance involves working with primary intended users of an evaluation to understand and agree on everything from the core evaluation questions to the evaluation design, and ultimately the analytical framework for making sense of data and rendering evaluative judgments (Patton, 2012). Discussing attribution versus contribution is part of arriving at a shared understanding of the evaluation's analytical framework.

Attribution is a central issue in evaluation: Can observed outcomes be directly attributed to the program? However, under conditions of complexity, with multiple interdependent variables interacting (often in the form of simultaneous program interventions overlapping and collaborating), traditional attribution analysis gives way to context-sensitive *contribution analysis* (Mayne, 2008;

Corresponding author:

Michael Quinn Patton, *Utilization-Focused Evaluation*, 740 Mississippi River Blvd So, Suite 21-E Saint Paul, MN 55116-1037, USA.

Email: mqpatton@prodigy.net

Patton, 2011: 244). When, as is typical, many causal factors contribute to the outcomes associated with an intervention, those who must make judgments about the merit, worth, and significance of that intervention still want to know the degree to which the intervention itself made a difference. The evaluation question becomes more nuanced: To what extent and in what ways has the program contributed to observed outcomes?

Observed outcomes does not just mean intended outcomes. A program or intervention may contribute to both intended and unintended outcomes. For example, *Outcome Harvesting* (Wilson-Grau and Britt, 2012) is a method of identifying, formulating, verifying and making sense of the full range of outcomes that may be associated with an intervention, including individual outcomes or groups of outcomes. The broader questions become: What outcomes have emerged, both intended and unintended? In what ways and to what extent have intervention activities contributed to those outcomes?

Evaluators bring evidence to bear on these questions. Evidence requires interpretation. What kinds of evidence with what degree of certainty is needed to render a judgment about contribution? And who will render this judgment? This is where utilization-focused evaluation enters the interpretive scene. The primary intended users of the evaluation must decide what kind of evidence they need and must weigh the evidence to determine the extent to which the preponderance of evidence supports a conclusion of contribution. In essence, they determine the degree of contribution: great, substantial, some, little, or none – or whatever adjectives *they choose* to express their interpretation. The evaluator's role under this scenario is to organize the evidence and facilitate the contribution interpretation. The evaluator will render his or her own independent judgment when requested to do so by primary intended users. Ultimately, evaluation users must own the judgment if they are to have sufficient confidence in the findings, and their interpretation of the findings, to take evidence-informed action. That said, the extent to which the evaluator, or evaluation team, renders an independent judgment about issues like attribution and contribution, or works collaboratively with primary intended users to arrive at a consensus on these matters, varies again by evaluation purpose and political context. The evaluator's role and relationship with primary intended users must be negotiated and made explicit in each evaluation.

In this article I shall elaborate the connection between utilization-focused evaluation and contribution analysis. In particular, I will examine the process of working with primary intended users in rendering judgments about the extent to which the preponderance of evidence supports a meaningful and useful conclusion about degree to which an intervention has affected observed outcomes and impacts. Two in-depth examples will illustrate this process. The first example concerns an evaluation of the causes of death of railroad employees killed in switching yard accidents for the purpose of generating accident-prevention safety procedures for the railroad industry throughout the USA. The second example is the evaluation of the contribution of the Paris Declaration to improving international aid effectiveness. These two quite different evaluation cases shared a utilization-focused evaluation process at the heart of which was contribution analysis. The ultimate value and utility of both evaluations depended, I shall argue, on the primary intended users understanding the implications of and buying into a contribution analysis framework. *Understanding* meant appreciating both the strengths and limitations of contribution analysis. *Buying into* meant that they were prepared to engage seriously and thoughtfully with findings presented through a contribution lens and draw their own conclusions about the strength of the evidence and the degree of contribution supported by the evidence.

Utilization-focused evaluation

Utilization-focused evaluation begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any

evaluation with careful consideration of how everything that is done, *from beginning to end*, will affect use. Use concerns how real people in the real world apply evaluation findings and experience the evaluation process. Therefore, the *focus* in utilization-focused evaluation is on *intended use by intended users*. Intended users refers to any and all stakeholders with a specific interest in using the evaluation outcomes to improve a program or make decisions about an intervention. Primary intended users can include program staff, administrators, participants (intended beneficiaries), community members, policy makers, and funders. How many and which intended users are directly involved in the evaluation process varies greatly depending on the evaluation's purpose, resources, timing, methods, and capacity building agenda. In utilization-focused evaluations, the evaluator does not design the evaluation in isolation from primary intended users. Since no evaluation can be value-free, utilization-focused evaluation answers the question of whose values will frame the evaluation by working with clearly identified, primary intended users who have responsibility to prioritize evaluation questions, identify appropriate methods, apply evaluation findings and implement recommendations.

Utilization-focused evaluation is highly personal and situational. The evaluation facilitator develops a working relationship with intended users to help them determine what kind of evaluation they need. This requires negotiation in which the evaluator offers a menu of possibilities within the framework of established evaluation standards and principles. Utilization-focused evaluation does not advocate any particular evaluation content, model, method, theory, or even use. Rather, it is a process for helping primary intended users select the most appropriate content, model, methods, theory, and uses for their particular situation. *Situational responsiveness* guides the interactive process between evaluator and primary intended users. This means that the interactions between the evaluator and the primary intended users focus on fitting the evaluation to the particular situation with special sensitivity to context. A utilization-focused evaluation can include any evaluative purpose (formative, summative, developmental), any kind of data (quantitative, qualitative, mixed), any kind of design (e.g. naturalistic, experimental), and any kind of focus (processes, outcomes, impacts, costs, and cost-benefit, among many possibilities). *Utilization-focused evaluation is a process for making decisions about these issues in collaboration with an identified group of primary users focusing on their intended uses of evaluation.*

A psychology of use undergirds and informs utilization-focused evaluation: intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings; they are more likely to understand and feel ownership if they have been actively involved; by actively involving primary intended users, the evaluator is training users in use, preparing the groundwork for use, and reinforcing the intended utility of the evaluation every step along the way (Patton, 2012).

Contribution analysis and utilization-focused evaluation

In working with primary intended users, it can be quite useful to distinguish between attribution analysis and contribution analysis. Mayne (2001, 2008, 2011) suggests beginning by distinguishing attribution questions from contribution questions. A utilization-focused evaluator can use these distinctions to help primary intended users determine what kind of inquiry and analysis will be appropriate and possible for the evaluation situation in which they find themselves:

Traditional causality questions (attribution)

- Has the program caused the outcome?
- To what extent has the program caused the outcome?
- How much of the outcome is caused by the program?

Contribution questions

- Has the program made a difference? That is, has the program made an important contribution to the observed result? Has the program influenced the observed result?
- How much of a difference has the program made? How much of a contribution? (Mayne, 2011: 64)

Contribution analysis is especially appropriate where there are multiple projects and partners working toward the same outcomes, and where the ultimate impacts occur over long time periods influenced by several cumulative outputs and outcomes over time. It is important to discuss with primary intended users the differences between attribution and contribution early in the evaluation design process, before data collection, so that intended users participate in framing the core evaluation questions and understand what kinds of judgments will be possible given the complexity of the intervention environment and the kinds of data that can be collected. The evaluator can help primary intended users anticipate what level of evidence they will need to credibly answer their central questions to their own satisfaction, including understanding the challenges of establishing causality.

An example of a utilization-focused evaluation based on contribution analysis was an evaluation of an advocacy campaign funded by several philanthropic foundations aimed at influencing a US Supreme Court decision (Patton, 2008). The collaborating foundations committed over US\$2 million to a focused advocacy effort within a window of nine months to potentially influence the Court. The evaluation case study examined the following question: *To what extent, if at all, did the final-push campaign influence the Supreme Court's decision?* The foundation's staff and Board wanted to know if funds spent on judicial advocacy made any discernible difference. If so, they wanted the evaluation to identify lessons that might influence such efforts in the future – and thereby guide foundation grant-making. In designing the evaluation, we needed to be clear with the primary intended users that many factors influence a Court decision and that, beyond the official judgment rendered by the Supreme Court about what influenced and informed its judgment, all other evidence would be indirect, at best, based on interviews with key knowledgeable and careful review of relevant documentation.

In working with the senior foundation staff and board representatives to understand contribution analysis and consider how to make conjectures about what influenced the Court's decision, a useful analogy was the judicial model itself. In any case before a court, a judge and/or jury hears evidence, often conflicting and confused evidence, and sorts it out as best they can to reach an informed and hopefully fair judgment based on the cumulative evidence. This was the standard we proposed applying to the advocacy campaign evaluation. We discussed with the primary intended users the distinction between attribution and contribution. Attribution, we explained, involves testing whether *a* causes *b*. In pharmaceutical research, for example, randomized control trials are conducted comparing a drug with a placebo to establish whether the relief of symptoms can be directly attributed to the drug. This straightforward notion of cause-effect works well for simple, bounded, and linear problems. However, a retrospective evaluation of what influenced a Supreme Court decision would not permit attribution. We could, however, provide evidence from multiple sources and documents to support a contribution analysis. Contribution analysis works well for understanding and interpreting results in complex systems where a variety of factors and variables interact dynamically within the interconnected and interdependent parts of the open system. Contribution analysis focuses on identifying *likely influences*. Contribution analysis, like detective work, requires connecting the dots between what was done and what resulted, examining a

multitude of interacting variables and factors, and considering alternative explanations and hypotheses, so that in the end, those involved can reach a reasonable judgment based on the preponderance of evidence. It was critical that the primary intended users understand what was possible and agree to a design that would support a contribution analysis. We were also asked to render our own judgment, which we did. We concluded:

Based on a thorough review of the campaign's activities, interviews with key informants and key knowledgeable, and careful analysis of the Supreme Court decision, we conclude that: **The coordinated final-push campaign contributed significantly to the Court's decision.** (Patton, 2008: 1)

With that example as context, let me turn now to the evaluation of the causes of railroad switching yard accidents.

Systematic, rigorous contribution analysis: An example

One of the most rigorous contribution analysis evaluation processes I've worked with was a *Switching Operations Fatality Analysis* (SOFA, 2010). Fifty-five railroad employees died in switching yard accidents from 2005 to 2010. To analyze the causes of these deaths, a multi-stakeholder SOFA Working Group was established made up of representatives from the railroad industry, labor unions, locomotive engineers, and federal regulators. These stakeholder groups have traditionally been suspicious of each other and often in conflict over regulations and procedures. However, they put aside those conflicts to rigorously investigate the causes of switching operations fatalities. They carefully reviewed every case, coding a variety of variables related to conditions, contributing factors, and kinds of operations involved. They then looked for patterns in qualitative case data and correlations in the quantitative cross-case data. They spent 3–4 hours coding each case and hours analyzing patterns across cases. They found that fatalities happen for a reason. Accidents are not random occurrences, unfortunate events, or just plain bad luck. The risks to employees engaged in switching operations are real, ever present, and preventable. The data showed patterns in why switching fatalities occur. Knowledge about the causes of railroad accidents have accumulated through this rigorous analysis over time and across cases. The first SOFA Working Group report was released in 1999 based on analysis of 76 railroad fatal accident case files. The 2004 report examined the 48 switching fatalities that had occurred since the first report. These analyses identified the 5 *LIFESAVERS* and 10 Special Switching Hazards that contribute to fatal accidents. The 2010 report presented the latest findings on the causes of deadly accidents.

On 25 February 2010, over 50 senior industry executives, labor leaders, and Federal Railroad Administration staff convened to discuss draft findings and their implications. The meeting opened with a moment of silence in memory of the 179 lives lost in switching operations. The group then reviewed how the SOFA Working Group arrived at its findings and determined that the results were credible, accurate and important. In making that determination, understanding contribution analysis, both what it could provide and its limitations, was critical. These key stakeholders needed the opportunity to discuss together how to make sense of the data. In small groups, they discussed patterns in the findings and identified potential preventive initiatives. The final SOFA report incorporated the reactions at this summit meeting and invited those in the railroad industry at all levels to engage with the findings and take actions to prevent fatalities. The goal remains zero deaths.

The SOFA Working Group analysis process was evaluated by a team of external, independent evaluation professionals (Bonnet et al., 2009). The evaluation included examining the quality of the analysis process, the validity of findings, what lessons have been learned, and what has changed

in the industry over the last decade that affects interpretation and use of the new findings. The independent evaluation concluded that the SOFA analysis process was systematic, rigorous, comprehensive, and objective. The findings were deemed valid and significant. At the core of the meta-evaluation was the conclusion that contribution analysis was the appropriate framework for making sense of the data.

Evaluation of the Paris Declaration

The Paris Declaration on Aid Effectiveness, endorsed in 2005, is a landmark international agreement and program of reform – the culmination of several decades of attempts to improve the quality of aid and its impacts on development. The Paris Declaration asserts five principles and 56 commitments. The Phase 2 Evaluation Report was an independent global evaluation of these efforts to improve the effectiveness of international aid, especially since 2005 (Wood et al., 2011; please note that all quotations and references below are to this report). The evaluation findings, conclusions and recommendations are aimed at government ministers, legislators, aid administrators and other specialized users, as well as to wider publics with an interest in development and aid. The evaluation aimed at both accountability and future improvements. The opening of the report asserts that the ‘underlying stakes are huge: better lives for billions of people (reflected in the approaching Millennium Development Goals for 2015); hundreds of billions of dollars expended; vital international relationships; and growing demands to see results from development aid.’

The evaluation included more than 50 studies in 22 partner countries and across 18 donor agencies, as well as several studies on special themes. It was conducted over four years, in two phases between 2007 and 2011. The overall results were distilled in a Synthesis Report organized around three main questions:

1. What are the factors that have shaped and limited the implementation of the Declaration reforms and their effects? (Factors contributing to implementation)
2. What improvements have been made in aid effectiveness as targeted in the Declaration? (Contributions to Aid Effectiveness)
3. What contributions have improvements in aid effectiveness made to sustainable development results? (Contributions to Development Results)

Built into these questions was a *contributions analysis* framework. The complexity of the international aid environment and development processes meant that no simple causal model would be either valid or useful. This point was emphasized in a thematic study of the Paris Declaration, aid effectiveness, and development effectiveness led by distinguished evaluator Elliot Stern that established the theoretical and methodological foundation for the Phase 2 evaluation. The thematic study noted that:

As the implementation of the PD [Paris Declaration] appears to be contextualized and influenced by specific ‘starting conditions’ and histories it is likely to be highly varied. The interaction and sequencing of factors are likely to change over time, and ‘two-way’ causalities are possible. This adds to the diversity that may arise from differences in interpretation and the co-existence of many other local policies and international programmes . . . Simple ‘logic models’ will not be easily applied. (Stern, 2008: viii)

The thematic study laid the foundation for contribution analysis as the framework for the Paris Declaration evaluation:

It will be difficult to separate out PD effects from that of other policies and initiatives – there will be problems of attribution; and contributions will be more likely to be exemplified in a narrative rather than measured, especially in the shorter term. (Stern, 2008: 43)

That is the approach the evaluation team adopted. The evaluation's contribution analyses examined whether and how the commitments, actors and incentives brought together by the Declaration and the Accra Agenda for Action have made a discernible and plausible difference. In so doing, the evaluation traced the logic of how the Declaration is supposed to work and, based on the evidence collected through multiple methods, tells the story of the complex pathways from development hopes to actual results. The evaluation both highlighted the complex and powerful influences at work in the development process as well as illuminating the realistic limitations and challenges of aid. The contribution analysis portrayed development as a *journey*. Using that metaphor, the methodology focused on assessing the *direction of travel* on each key principle and outcome, and the *pace and distance* travelled so far. Multiple sources of evidence and techniques were used to validate answers and reach judgments.

In so doing, the evaluators knew they would have to address 'the thorny issues of *causality* and *attribution*' (Wood et al., 2011: 204). Given that the Paris Declaration implementation is a multidimensional, multi-level process, affected by many factors, the developmental journey in any given context can change direction, emphasis, and pace at different times and in response to many different influences. Thus, the evaluators sought a framework that could describe any relationship between Declaration implementation and accelerated development results while taking into account the following realities:

1. much of the change process toward aid effectiveness started in many countries *before* the Declaration was formally signed and implemented (i.e. pre-2005);
2. aid is only one of many potential contributory factors to development; and
3. other features of a context (governance/political changes/economic crisis/institutional capacity, etc.) can have far greater effects on development results than aid reform or aid as a whole.

Given these realities, the evaluation team explicitly adopted a contribution analysis framework. The report's technical appendix explains:

Traditional linear approaches to evaluation which would aim to causally 'attribute' change to the Declaration were consequently not appropriate here. As the main Synthesis Report makes clear, a political statement cannot by itself cause change; rather, what the Evaluation aimed to research is whether the operational commitments, relevant actors and motivational elements that it helped bring together from many sources have actually contributed to the intended improvements. It was therefore agreed with the International Reference Group that, due to the diffuse nature of the Evaluation object in this case, the use of any simple 'counterfactual' in assessing progress would not be viable methodologically. Attempting to draw linear lines of 'attribution' between Declaration implementation and development results (in sectors or in the form of poverty indicators for example) would be even more flawed as an approach.

Consequently, the Core Team opted to steer a course in favor of '*contribution*' rather than direct attribution, drawing on the work of Mayne (2001) and others and also applied in recent multi-country studies of e.g. Sector Budget Support. Teams were asked to assess the '*plausible contributions*' of the Declaration in their contexts to development results; to clearly evidence any changes and connections observed; and to state as explicitly as possible any other plausible explanations.

To support the application of this approach, specific guidance was produced and circulated. The guidance proposed that teams trace back from the *development results achieved (Question 3)* through to the *changes in the way aid has been delivered since the Declaration was enacted (Question 2)* through to: *the context and the extent of implementation of Declaration principles (Question 1)*, and to explore the links and connections between these processes. (pp. 204–5)

The International Reference Group's endorsement of the contribution analysis framework

The preceding section explains why the evaluation team adopted a contribution analysis framework, how they sequenced the questions and formulated the case studies to support such an analysis. However, for that approach to have credibility and utility, the primary stakeholders involved in advising the evaluation had to understand and endorse contribution analysis. The International Reference Group (IRG) that advised the evaluation included participants from the 22 partner countries, donors, multilateral organizations, civil society representatives, and, chiefly, members of the OECD/DAC Network on Development Evaluation as well as members of the OECD/DAC Working Party on Aid Effectiveness. The IRG was co-chaired by representatives from Malawi and Sweden and met four times during Phase 2 to provide guidance to the evaluation process. I observed the 3rd and 4th meetings of the IRG in Indonesia and Copenhagen, respectively, as part of a meta-evaluation of the Paris Declaration Evaluation (Patton and Gornick, 2011). The IRG members reviewed and commented on the evaluation framework and the inception report, emerging findings and the draft Final Report. The IRG advised on evaluation design, data collection, analysis, and reporting. Such stakeholder engagement can be difficult, challenging, expensive, controversial, conflict-laden, and undermining, and will be all those things to an extent that harms the evaluation's credibility unless the process is carefully planned and facilitated. The more diverse the group, the greater will be the risks. The meta-evaluation concluded that the IRG process was carefully planned and executed to provide opportunities for input into all aspects of the evaluation while maintaining the independence of the evaluation team, which had final responsibility and authority (Patton and Gornick, 2011).

In the International reference Group meetings, the collaborative process allowed input on and reactions to all aspects of the evaluation design and synthesis report. In the December 2010 IRG meeting in Bali, IRG members had an opportunity to react to emerging findings. The Core Evaluation Team responded systematically to how they would incorporate the feedback they received, including presenting appropriate changes in language to reflect the responses and reactions offered. This process continued and intensified at the IRG meeting in Copenhagen in April 2011. Every chapter of the draft report was reviewed by small groups. The groups reported their reactions in plenary. The Core Evaluation Team responded to the feedback they were offered, being clear about what changes they could make based on the evidence, and always expressing appreciation for the group's input and reaffirming that, in the end, the Core Evaluation Team would determine and own the findings, conclusions, and recommendations.

Throughout these interactions and deliberations the nature of the contribution analysis was a prominent issue of discussion. There was in-depth engagement around what conclusions the evidence supported, what it meant to make plausible connections, and the relative strength of various contributory factors. The IRG discussed at length the evaluation's limitations with particular attention to issues of attribution versus contribution. Recognizing and understanding what the evaluation could and could not do proved important in the decision to frame the evaluation methodology as contribution analysis. The implications of this framing became especially evident

in the evaluation's analysis of the Declaration's contributions to development results and the IRG discussion of those findings.

Contributions to development results

The Evaluation concentrated on assessing the possible contributions of aid reforms to sustainable development in four areas: in specific sectors (particularly in health, the common study sector for the country evaluations); in giving priority to the needs of the poorest; in strengthening institutional capacities and social capital; and in improving the mix of aid modalities. The evaluation concluded:

Significant positive contributions can be traced, particularly in the case studies in the health sector, to more focused aid efforts and better development results. The pathways for these contributions are indirect but clear. In other areas assessed – aid and aid reform have been able to make little difference to development gains in the face of powerful obstacles. A number of the gains made are likely to be sustainable, while others remain fragile. (pp. xiv–xv)

Throughout the report, the evaluation's analysis and conclusions are careful to describe contributions rather than direct causes. For example:

A strong cross section of the country evaluations found evidence that Declaration type measures, launched either before or since 2005, but reinforced since then, have contributed to more focused, efficient and collaborative aid efforts, particularly at the sectoral level. These evaluations then found plausible evidence that those efforts had already contributed to better development results, with good prospects of being sustainable. The strongest evidence of this effect is in the health sector, examined in most depth in the country evaluations. Beyond this 'tracer' sector, this Evaluation does not have sufficient evidence to track contributions of aid reforms to wider development results such as accelerating achievement of the other Millennium Development Goals.

Although insufficient capacity remains a formidable obstacle in many countries and aid could help more than it does, there is evidence that aid and aid reform have made at least some contributions to the long-term strengthening both of institutional capacities for development and of social capital.

On the whole, there has been little progress in most countries in giving greater priority to the needs of the poorest people, particularly women and girls. However, there is evidence of some positive contributions by aid and some value added by reforms and Declaration-style operations since 2000–05. This disconnect drives home the essential precondition of a powerful and sustained national commitment to change. Without this in place, aid and aid reforms are limited in their capacity to address entrenched inequalities. A wider range of options and innovations with aid modalities, particularly more joint donor support at the sectoral level, has improved actual or potential contributions to development results in half the evaluation countries since 2000–05. However, the Evaluation shows that no single modality (e.g. budget or sector support, programs or projects) will automatically produce better development results, and a mix of aid modalities has continued to make sense for all partner countries and donors. (p. xv)

The contribution analysis framework is most evident in addressing the third core question: Has the implementation of Paris Declaration strengthened the contribution of aid to sustainable development results? If so, how? This part of the evaluation moved from analyzing changes in aid effectiveness to analyzing the evidence on whether the Declaration campaign had yet contributed to results. The conclusion:

None of the evaluations finds an immediate and direct connection between the Declaration campaign and development results achieved, but there is evidence in a solid majority of the reports that it has made at least some *plausible contributions* to better results through the pathways of the reforms traced. The analysis follows the agreed framework for assessing the possible contributions of aid reforms to sustainable development results. The contributions of aid – and in turn any reforms – are dependent on the development progress achieved in a country. (p. 43)

The report explains the challenges of using a contribution approach, noting that all of the evaluators involved in the case study countries knew that tracing development effects would be difficult:

Before crediting the Declaration reforms with any contributions, the evaluations have taken into account the influence of other factors in improved aid contributions and development results, and weighed alternative explanations including the possibility that the results might have occurred anyway in the absence of these contributions. *Thus they applied the basic rule of the 'contribution analysis' approach to evaluation.* (p. 43; emphasis added)

The evaluation design never anticipated that better aid practices, if achieved, could directly or rapidly lead to better development results – demonstrable improvements in the lives of people in partner countries, particularly the poor and vulnerable – in a five year period. Since many other factors are usually more important than aid in determining these results, the country evaluations looked for:

- first, evidence of development changes; then
- plausible evidence as to whether aid had contributed to such changes; and
- if so, plausible evidence that aid reforms might have strengthened the aid contribution.

Based on these evidentiary steps, the report concluded that:

In fact the findings exceed the very modest expectations of contributions to results in this short period. A strong cross section of the country evaluations found evidence that Declaration type measures, either launched before or since 2005, but usually reinforced since then, have contributed to more focused efforts, particularly at the sectoral level. These evaluations then found evidence that those efforts had already contributed to better development results, with good prospects of being sustainable. These effects were found mainly in the health sector, which had been selected for more detailed examination in almost all the country evaluations. (p. 207)

The IRG meetings involved diverse stakeholders in reviewing these and related conclusions. The constant theme throughout centered on generating conclusions and rendering judgments about Paris Declaration contributions. It was critical to the credibility of the evaluation that IRG members understood the strengths and limitations of contribution analysis, not only so that they could determine for themselves the meaningfulness and implications of the evaluation findings, but so that they could communicate the evaluation's framework and findings to the constituencies they represented, which included political leaders in their home countries and the organizations that designated them as IRG members.

Capacity-building to conduct contribution analysis at the country level

Another important element of contribution analysis was building the capacity of evaluators in the case study partner countries to undertake contribution analysis. The Core Evaluation Team ran

workshops for country level evaluators in which the nature of contribution analysis was a featured element. As the country evaluations were being conducted, the Core Evaluation Team kept in close contact with country teams and products emerging from the process. Core Team members had ongoing insight into methods being applied (mostly standard qualitative and quantitative techniques), but also areas of challenge that arose, mostly around applying contribution analysis. To help them undertake the country case study contribution analysis, teams were provided with a list of potential ‘mechanisms of change’ to consider as potential tools for explaining and/or categorizing documented changes. The technical annex reports that the majority of evaluation teams found the results logic of contribution analysis very challenging:

for several, the shift from direct attribution to contribution analysis required new capabilities and understanding. There were some teams for whom intensive support was clearly a pre-requisite for ensuring the completion of their evaluation to provide full input for the Synthesis level. The support provided, in the various forms of guidance, detailed explanations, the provision of examples around issues and the working terms of the Evaluation, plus comments on inception and draft reports, appeared well received by teams. Its value was noticeable when reports were later analyzed at Emerging Findings stage. (p. 218)

The technical annex of the Paris Declaration concludes with this methodological observation:

The challenges of mapping the contributions of a political Declaration to changed performance and results cannot be underestimated – [but] the value of contribution analysis is confirmed; [and] a clear framework for exploring and explaining pathways of contributions is essential as guidance for evaluation teams. (p. 218)

Rigorous thinking supersedes rigorous methods

Both the SOFA working group and the Paris Declaration evaluation teams used multiple methods, drew on diverse sources and types of data, carefully examined and questioned assumptions, and thoughtfully triangulated findings as part of their respective contribution analyses. Rigorous thinking was more important than reliance on methodological rigor. This is a critical distinction. In concern about research rigor, we risk emphasizing adherence to method over rigorous inquiry and critical thinking. Contribution analysis depends on critical thinking. In considering what this means, it is helpful to review conclusions about what constitutes rigor in information analysis done by intelligence analysts. As a result of the fiasco of faulty military intelligence leading up to and used to justify the invasion of Iraq, leaders in the intelligence community undertook a comprehensive review to reconsider what rigorous analysis means. Their findings are relevant to contribution analysis, I believe. Indeed, I have shared what follows with primary intended users in designing an evaluation to help them appreciate the importance of rigor and think about what it means.

They began by asking: ‘Given the many contexts in which *information analysis* occurs, what does it mean to be rigorous? Moreover, how do analysts go about achieving acceptable levels of rigor in their analysis processes? And what might help them to be more rigorous?’

Their work led to a model of analytical rigor that frames the concept ‘rigor’ as an emergent multi-attribute measure of sufficiency (‘Has the analyst done enough?’) rather than as it is commonly framed, as a measure of process adherence (‘How rigidly has the analyst followed a particular method?’). They identified eight attributes of rigorous analysis that contribute to the assessment of the quality and the sufficiency of an analysis (Woods, 2007).

Attributes of the rigor metric

1. *Hypothesis Exploration* describes the extent to which multiple hypotheses were considered in explaining data. In a low-rigor process there is minimal weighing of alternatives. A high-rigor process, in contrast, involves broadening of the hypothesis set beyond an initial framing and incorporating multiple perspectives to identify the best, most probable explanations.
2. *Information Search* relates to the depth and breadth of the search process used in collecting data. A low-rigor analysis process does not go beyond routine and readily available data sources, whereas a high-rigor process attempts to exhaustively explore all data potentially available in the relevant sample space.
3. *Information Validation* details the level at which information sources are corroborated and cross-validated. In a low-rigor process little effort is made to use converging evidence to verify source accuracy, while a high-rigor process includes a systematic approach for verifying information and, when possible, ensures the use of sources closest to the areas of interest.
4. *Stance Analysis* is the evaluation of data with the goal of identifying the stance or perspective of the source and placing it into a broader context of understanding. At the low-rigor level an analyst may notice a clear bias in a source, while a high-rigor process involves research into source backgrounds with the intent of gaining a more subtle understanding of how their perspective might influence their stance toward analysis-relevant issues.
5. *Sensitivity Analysis* considers the extent to which the analyst considers and understands the assumptions and limitations of their analysis. In a low-rigor process, explanations seem appropriate and valid on a surface level. In a high-rigor process the analyst employs a strategy to consider the strength of explanations if individual supporting sources were to prove invalid.
6. *Specialist Collaboration* describes the degree to which an analyst incorporates the perspectives of domain experts into their assessments. In a low-rigor process little effort is made to seek out such expertise, while in a high-rigor process the analyst has talked to, or may be, a leading expert in the key content areas of the analysis.
7. *Information Synthesis* refers to how far beyond simply collecting and listing data an analyst went in their process. In the low rigor process an analyst simply compiles the relevant information in a unified form, whereas a high-rigor process has extracted and integrated information with a thorough consideration of diverse interpretations of relevant data.
8. *Explanation Critique* is a different form of collaboration that captures how many different perspectives were incorporated in examining the primary hypotheses. In a low-rigor process, there is little use of other analysts to give input on explanation quality. In a high-rigor process peers and experts have examined the chain of reasoning and explicitly identified which inferences stronger and which are weaker. (Zelik et al., 2007).

In sharing this framework with a variety of primary intended users, I find that they can understand the dimensions of rigor identified here and, especially relevant to this discussion, they can appreciate the importance of rigorous thinking to engage in contribution analysis. (Note of clarification in response to a reviewer's question: I did not use this framework in the examples presented here because I did not know about it when working with the cases discussed here.)

Engaging stakeholders in critical thinking within a contributions analysis framework

What stands out to me about these criteria for rigor is the emphasis on engaging multiple perspectives, deep questioning, and critical thinking. There is the danger that as stakeholders learn about the nonlinear dynamics of complex systems and come to value contribution analysis, they will be inclined to always find some kind of linkage between implemented activities and desired outcomes. One reviewer noted that in all three examples, the evaluation concluded that a significant contribution was made. The reviewer then went on to ask whether it is likely that a contribution analysis will always find a contribution of some kind. In essence, the concern is that treating contribution as the criterion (rather than direct attribution) is so weak that a finding of *no contribution* is highly unlikely. This is an appropriate concern. More experience with contribution analysis will be needed to answer this question. My sense is that the answer will depend on the rigor of the contribution analysis, which is why I've included the rigor attribute model here as an analytical framework for enhancing rigor.

Contribution analysis involves a particular logic of what may be thought of as *connecting the dots* between activities and observed outcomes in complex dynamic environments. Simple linear causality is an overly narrow way of understanding the complex interconnections in complex systems where nonlinear interactions, iterative feedback loops, and interdependencies prevail. Attribution analysis tends to focus on direct, verifiable causality. Contribution analysis broadens the questions to include a full continuum and range of degrees of connection and relationship between the activities undertaken and the results observed. This often involves forensic logic and abductive inference (Patton, 2011: 284–7).

The credibility and utility of contribution analysis ultimately depends upon both the quality of evidence collected and the quality of thinking applied in making sense of that evidence. It is not enough for evaluators to understand and engage in rigorous contribution analysis. The central point of this article is that primary intended users involved in the evaluation must also have an opportunity to learn about, come to understand and appreciate, and engage in contribution analysis. In that way contribution analysis contributes to evaluation use.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- Bonnet D, Ranney J, Snow J, Coplen M and Patton MQ (2009) Evaluation of the Switching Operations Fatality Analysis 2010 Working Groups Processes. Cambridge, MA: John A. Volpe National Transportation Systems Center. Research and Special Programs Administration, Report Number: DOT/FRA/ORD-09/23.
- Funnell SC and Rogers PJ (2011) *Purposeful Program Theory: Making Effective Use of Theories of Change and Logic Models*. San Francisco, CA: Jossey-Bass.
- Mayne J (2001) Addressing attribution through contribution analysis: using performance measures sensibly. *Canadian Journal of Program Evaluation* 16(1): 1–24.
- Mayne J (2008) *Contribution Analysis: An Approach to Exploring Cause and Effect*. ILAC Brief No. 16: Rome: The Institutional Learning and Change Initiative.
- Mayne J (2011) Contribution analysis: addressing cause and effect. In: Schwartz R, Forss K and Marra M (eds) *Evaluating the Complex*. New Brunswick, NJ: Transaction Publishers, 53–96.
- Patton MQ (2008) Advocacy impact evaluation. *Journal of MultiDisciplinary Evaluation* 5(9): 1–10.

- Patton MQ (2011) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford.
- Patton MQ (2012) *Essentials of Utilization-Focused Evaluation*. Thousand Oaks, CA: SAGE.
- Patton M and Gornick J (2011) *Evaluation of the Phase 2 Evaluation of the Paris Declaration*. Copenhagen. URL: <http://www.oecd.org/dataoecd/22/29/48461537.pdf>
- SOFA Working Group (2010) *Switching Operations Fatality Analysis*. Washington, DC: SOFA Working Group.
- Stern ED with Altinger L, Feinstein O, Marañón M, Ruegenberg D, Schulz N and Nielsen NS (2008) *Thematic Study on The Paris Declaration, Aid Effectiveness and Development Effectiveness*. Madrid: DaRa Development Assistance Research Associates and Copenhagen: Ministry of Foreign Affairs. URL: <http://www.oecd.org/dataoecd/59/28/41807824.pdf>
- Wilson-Grau R and Britt H (2012) *Outcome Harvesting, 2012*. New York: Ford Foundation. URL: <http://www.fordfoundation.org>
- Wood B, Betts J, Etta F, Gayfer J, Kabell D, Ngwira N, Sagasti F and Samaranayake M (2011) *The Evaluation of the Paris Declaration, Final Report*. Copenhagen: OECD. URL: http://www.oecd.org/document/60/0,3746,en_21571361_34047972_38242748_1_1_1_1,00.html
- Woods DD (2007) *Revealing Analytical Rigor: A Strategy for Creating Insight into the Information Analysis Process*. URL: <http://academic.research.microsoft.com/Paper/13513837.aspx>
- Zelik D, Patterson ES and Woods DD (2007) *Understanding Rigor in Information Analysis*. URL: http://csel.eng.ohio-state.edu/zelik/research/Rigor_files/ZelikPattersonWoods_UnderstandingRigor_1.pdf

Michael Quinn Patton is an independent evaluation consultant based in the United States. He is former president of the American evaluation Association. He is author of *Utilization-Focused Evaluation* (4th edn, 2008), *Essentials of Utilization-Focused Evaluation* (2012), and *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use* (2011).