

Making contribution analysis work: A practical framework for handling influencing factors and alternative explanations

Evaluation
18(3) 294–309
© The Author(s) 2012
Reprints and permission: sagepub.
co.uk/journalsPermissions.nav
DOI: 10.1177/1356389012450654
evi.sagepub.com


**Sebastian T. Lemire, Steffen Bohni Nielsen
and Line Dybdal**

Rambøll Management Consulting, Denmark

Abstract

This article examines the methodological strengths and weaknesses of contribution analysis. The authors contend that a salient characteristic of contribution analysis is its accounting for influencing factors and alternative explanations. We argue that contribution analysis in its current form needs further methodological and practical elaboration in this respect. Therefore, this article pays particular attention to how the methodology can identify and determine the extent of influencing factors and alternative explanations. The authors propose a corresponding framework – the Relevant Explanation Finder – which can help evaluators account for these elements and thereby create more credible contribution stories.

Keywords

alternative explanation, contribution analysis, influencing, mechanism, theory-based evaluation factor

Introduction

Contribution analysis (CA) has been presented and conceptually developed by John Mayne in a series of seminal papers (1999, 2001, 2008, 2009, 2011). During the course of its development, CA has made a transition from its original performance measurement context toward a new function, namely that of evaluating complex programs in complex settings (Mayne, 2011). While CA has undergone refinements of methodology, and even more notably of scope, its underlying logic and core steps have remained more or less unchanged: (i) elaborating the intervention's theory of

Corresponding author:

Sebastian T. Lemire, Rambøll Management Consulting, Hannemanns Allé 53, DK-2300 Copenhagen S, Denmark.
Email: slemire@ucla.edu

change; (ii) identifying key threats to the mechanisms of the theory of change; (iii) identifying other contributing factors; and (iv) testing the principal rival explanations (Dybdal et al., 2010).

In the context of evaluation studies, an evaluator using CA aims to infer ‘plausible association’ between the program and a set of relevant outcomes by means of systematic inquiry (Mayne, 1999: 5–7). This implies establishing evidence that supports the probability that the program – as presented in the form of an embedded theory of change – has contributed to a set of specified outcomes. According to Mayne, five criteria concerning the embedded theory of change must be met in order to infer ‘plausible association’ (2011: 7):

1. Plausibility: Is the theory of change plausible?
2. Implementation according to plan: Has the program been implemented with high fidelity?
3. Evidentiary confirmation of key elements: To what extent are the key elements of the theory of change confirmed by new or existing evidence?
4. Identification and examination of other influencing factors: To what extent have other influencing factors been identified and accounted for?
5. Disproof of alternative explanations: To what extent have the most relevant alternative explanations been disproved?

These collectively serve as the quality marker for the causal stories (contribution stories) elicited by evaluators using a CA-based methodology (Mayne, 2011: 7). Accordingly, the credibility of contribution stories rests on this foundation.

In spite of the conceptual advancement of CA, little attention has been awarded to *how* to examine influencing factors and alternative explanations (criteria four and five). Consequently, there is no operational framework for what criteria four and five entail in practice.

We find a practical elaboration of these criteria to be essential for three reasons. First, the focus on accounting for influencing factors and alternative explanations has been identified as one of the strong points of CA in relation to theory-based evaluations (Dybdal et al., 2010). Indeed, the methodological strength of CA is precisely its potential ability to systematically explore – without a counterfactual – the often complex and messy nature of programs by accounting for the influence of contextual factors and explanations. In order to deliver on this promise, evaluation practitioners applying CA in complex settings need an operational framework.

Second, the value of any evaluation ultimately depends on the credibility of the findings it generates. As such, the foundation of confidence in any design, method or analytical strategy that strives to be methodologically sound and aims to produce causal claims (by addressing either attribution or contribution) is contingent on how adequately it pays attention to the influence of competing factors and explanations. CA is no exception. Therefore, addressing this challenge is a necessary step in promoting CA as a credible evaluation methodology (Dybdal et al., 2010).

Third, earlier reviews argue that the practical application of CA suffers from its lack of an operational (and transparent) approach that accounts for influencing factors and alternative explanations (Dybdal et al., 2010). In its current form and application, CA fares well in terms of identifying the most salient influencing factors and explanations, but it still requires further practical elaboration in order to account adequately for the extent of their influence.¹ Accordingly, the evaluator is left in the dark as to how to translate into practice some of the key steps in building credible contribution stories. A practical framework for how to gauge the relative importance of these influencing factors and alternative explanations might serve to guide the evaluator through these important steps in CA.

Thus, the corresponding objective of this article is to address this challenge. The article is structured into three sections. In the first part, we briefly discuss key concepts concerning

mechanisms, influencing factors and alternative explanations. In the second part, we present, exemplify and discuss a proposed framework – the so-called ‘Relevant Explanation Finder’ (REF) – which is a practical framework that supports the systematic examination of both influencing factors and alternative explanations. We ground our discussion of the REF in an evaluation we conducted of a teaching assistant program in Denmark. In the third and final part, we briefly summarize our conclusions regarding our experiences with the REF and the implications for the further use of the framework in the context of CA.

Mechanisms, alternative explanations and influencing factors in evaluation

In the evaluation field, such terms as *mechanism*, *influencing factor* and *alternative explanation* have ambiguous meanings and are often confused. Conceptual clarity is therefore needed in order to create a systematic and operational methodology for CA. More specifically, we believe that the following key terms need to be clearly defined.

Mechanism

In the social sciences, the term ‘mechanism’ has been endowed with numerous definitions and ambiguous meanings (Hedström and Swedberg, 1996, 1998; Mayntz, 2004). In a recent article, Astbury and Leeuw (2010) reviewed the concept of mechanism in the context of the literature concerning evaluation. Drawing from a Realist tradition, they define mechanisms as ‘underlying entities, processes, or structures which operate in particular contexts to generate outcomes of interest’ (p. 368). This definition implies that mechanisms exhibit the following characteristics: (a) they are usually hidden; (b) they are sensitive to variations in context; and (c) they generate outcomes (p. 368). Inspired by Carol Weiss (1997), Astbury and Leeuw (2010) additionally propose that mechanisms can be viewed as ‘cognitive, social or affective responses’ that emerge, yet are distinct, from program activities (p. 367).

Astbury and Leeuw (2010) make an important distinction between program activities (e.g. a teaching aid resolving a classroom conflict), the ensuing mechanisms (e.g. the children feeling calm and focused on learning) and the generated outcomes (e.g. a better learning environment). This distinction emphasizes the key function of mechanisms as connectors between program activities and program outcomes and also suggests the importance of mechanisms in explaining how and why programs work (2010: 364). In the context of CA, the explanatory logic is just that: to explicate the constellation of mechanisms and contextual factors that constitute primary contributors to a particular set of outcomes. We explain how a program contributes to a set of outcomes by specifying the primary explanatory mechanism(s). The primary explanatory mechanism is the underlying mechanism of the intervention that accounts for and explains the observed outcomes (i.e. the primary explanatory mechanism in the contribution story). In some situations a primary explanation is hard, if not impossible, to predefine. As such, examining and delineating both primary and alternative explanations happens simultaneously (the case below is an example).

Alternative explanation

For us, an alternative explanation is best viewed as a competing mechanism that represents a plausible rival explanation for the observed outcome(s). In the context of CA, alternative explanations may potentially modify, or even undermine, the intended primary explanatory mechanism, and

hence affect the explanatory capacity of the contribution story (we will propose two types of alternative explanations in the framework below).

Influencing factor

Influencing factors are contextual conditions that might enable or impede mechanisms. While some influencing factors might activate or hinder the onset of mechanisms, others might enhance or dampen the ability of mechanisms to generate outcomes. The importance of identifying and accounting for influencing factors is well established in the evaluation literature, and some theorists (e.g. Pawson and Tilley, 1997), state that the recognition of contextual factors is essential for understanding how a causal mechanism works. Pawson et al. (2004) have suggested that contextual factors must be considered at different levels:

- The *individual* capacities of the key actors and stakeholders such as interests, attitudes, capabilities and the credibility of (for instance) professionals or beneficiaries.
- The *interpersonal* relationships required to support the intervention, such as lines of communication, management and administrative support, union agreements and professional contracts.
- The *institutional* setting in which the intervention is implemented, such as the culture, leadership etc. of the implementing body.
- The wider (*infra-structural* and welfare system, such as political support, the availability of funding resources etc. (Pawson and Greenhalgh, 2004: 65).

Influencing factors are therefore contextual conditions that are deemed to be particularly salient in shaping the behaviour of mechanisms.

In the remainder of this article, we will consider how this terminology might be integrated into a framework and used in the context of CA.

Making it work: A practical framework for handling influencing factors and alternative explanations

According to Mayne (2011), CA can primarily address attribution by providing answers to contribution-related questions such as: Has the intervention made a difference? How much of a difference has it made? Providing answers to these questions necessarily involves some kind of systematic approach to the examination of influencing factors and alternative explanations. In our practical application of CA, we have developed a framework – the REF – to account for both influencing factors and alternative explanations.

This section begins by outlining the REF and establishing the connections between the components of the framework and the previous discussion concerning terminology. We then present an example of the practical application of the REF and reflect on our experiences of using it during the course of a specific evaluation study.

Presenting the REF

The overall aim of the REF is to provide a guiding framework for the identification and examination of influencing factors and alternative explanations. As was mentioned above, this is a

necessary step in constructing a credible contribution story. Our approach builds on the notion that the consideration of context should take place throughout the evaluation, right from making the influencing factors and alternative explanations explicit in the early stages of the evaluation to gathering and weighing the evidence in favor of these factors and explanations when the contribution story is assembled toward the end of the evaluation.

The general intent of the REF is not to outline all the possible and external influencing factors and alternative explanations, but rather to formulate and present evidence regarding the most salient factors and explanations, seeking to show how the evidence might favour or reject specific constellations of these as primary contributors to the observed outcomes (Yin, 2003: 218). The REF not only aims to identify and list these rival explanations, but also to judge their relative importance in relation to the contribution story. Arguably, the choice concerning which influencing factors and alternative explanations should be examined is at least partly informed by subjective preferences. This only makes the need for a framework that enables the execution of a systematic and transparent prioritization of factors and explanations all the more pertinent.

It is important to note that the REF should not be viewed or employed as a mechanistic recipe or a set of rigid procedures that will automatically ensure the creation of credible contribution stories. Rather, it is an operational framework that provides a structure for enabling transparent and explicit decision-making regarding what explanations and influencing factors to focus on in the evaluative inquiry.

The framework, which is informed by Yin’s work on alternative explanations, consists of six components (see the columns in Table 1 below). The first two components are to be completed in the beginning of the evaluation, as part of framing the subsequent data collection and analysis. The remaining four components are intended for the analysis phase of the evaluation. The intended procedure for filling out and using the REF is otherwise unscripted. While some evaluators might use the framework as an internal tool to be filled out by the evaluation team, others might find it worthwhile to include clients or stakeholders.

Table 1 presents the format of the REF and, below, we consider each component of the REF in turn. The first column provides a description of the most salient influencing factors and alternative explanations, and is filled out at the beginning of the evaluation (several strategies for deciding on relevant explanations and factors are presented later in the article).

Table 1. The Relevant Explanation Finder.

1.	2.	3.	4.	5.				6.
Description	Type	Level	Identifiers	Degree of influence				Implication
				Certainty	Robustness	Range	Prevalence	Theoretical grounding
a.								
b.								
c.								
d.								
e.								

The second column focuses on the type of explanation or factor identified. Inspired by Yin (2003), we employ four categories in the REF:

1. *Primary explanation (mechanism)* – a mechanism identified and purported to be the target intervention mechanism that accounts for and explains the observed outcomes (i.e. the primary explanatory mechanism in the contribution story).
2. *Direct rival (mechanism)* – a mechanism, different from the target intervention mechanism, that accounts for and explains the observed outcomes (i.e. undermines the contribution story).
3. *Commingled rival (mechanism)* – other mechanisms, along with the target mechanism, that both contribute to and explain the observed outcomes (i.e. revise the contribution story).
4. *Implementation rival (factors)* – influencing factors in the implementation process, not substantive intervention mechanisms, that modify the outcomes (i.e. revise the contribution story).

Note that the two types of alternative explanations represent different challenges to the contribution story. While the direct rival undermines the core of the contribution story, the commingled rival merely modifies it.

The third column presents the various factor and explanation levels informed by the previous discussion regarding context. Using the categorizations outlined in the previous section, these are: (i) the individual level; (ii) the interpersonal level; (iii) the institutional level; and (iv) the (infra) structural level.

The fourth column concerns descriptions of the possible identifiers that offer proof or disproof of the existence of the various influencing factors and alternative explanations. Such identifiers may take the form of observable outcomes or patterns of outcomes that favor the existence of these influencing factors and explanations. It is imperative that a set of identifiers be specified and described for each of the influencing factors and explanations in advance of the data collection and analyses.

The fifth column summarizes the degree of influence of both the factors identified and the underlying mechanisms for the observed outcomes. There are five attributes to be considered in assessing the degree of influence:

1. *Certainty*: The degree to which the observed outcome pattern matches the one predicted by the factor or mechanism.²
2. *Robustness*: The degree to which the factor or mechanism is identified as a significant contributor across a broad range of data sources and data collection methods.
3. *Range*: The degree to which the factor or mechanism contributes to a broad range of the outcomes of interest.
4. *Prevalence*: The degree to which the factor or mechanism contributes to the outcomes of interest across a wide range of implementation environments and target groups (e.g. different implementation sites and / or types of intervention).
5. *Theoretical grounding*: The factor or mechanism is informed by theory (identifies existing theories of which it is an example) and is cast in specific terms (i.e. it is not vague).³

The sixth column contains conclusions regarding the implications of the factors and mechanisms for the contribution story. Here, the implications of the examination of influencing factors and alternative explanations are presented and considered as a component of the articulation of the final contribution story.

A more detailed description of the six steps and our experiences with the REF is presented in the section below, which concerns the application of the framework.

Applying the REF

This section takes as its basis our application of the framework in an evaluation of teaching assistant interventions in Danish public schools. A brief note on the evaluation is therefore appropriate. The evaluation was carried out in 2010–11, and was commissioned by the Danish Agency for Quality Assurance and Evaluation (Skolestyrelsen), a department located in the Ministry of Education. The teaching assistant interventions were implemented in two rounds in the course of two successive school years. The primary aim of the teaching assistant interventions was to provide the teacher with practical support in the classroom, and in effect to allow the teacher to focus on teaching, to support a better learning environment and to enhance the pupils’ ability to learn. In practice, the teaching assistants often combined several interconnected roles in the classroom. More specifically, there were three types of roles for the teaching assistant: the ‘practical assistant,’ where the teaching assistant primarily provides the teacher with practical support (e.g. copying, preparing the classroom, helping pupils with practical matters, etc.); the ‘pedagogical assistant,’ where the primary purpose of the teaching assistant is to provide the teacher with pedagogical support (e.g. by resolving conflicts among pupils, supporting relationship-building among pupils, etc.); and the ‘special pedagogical assistant,’ where the teaching assistant provides special pedagogical support to a select group of pupils with special needs. The mixing of roles in the actual implementation of the interventions presented a real challenge in terms of defining a strong primary explanation (a point that will be revisited below). As a primary explanation, the closest bidder was the ‘practical assistant,’ where the teaching assistant allows the teacher to focus on teaching and in effect creates a more stable and better learning environment. At the beginning of the evaluation, we developed a theory of change for the teaching assistant program (see Figure 1 below).

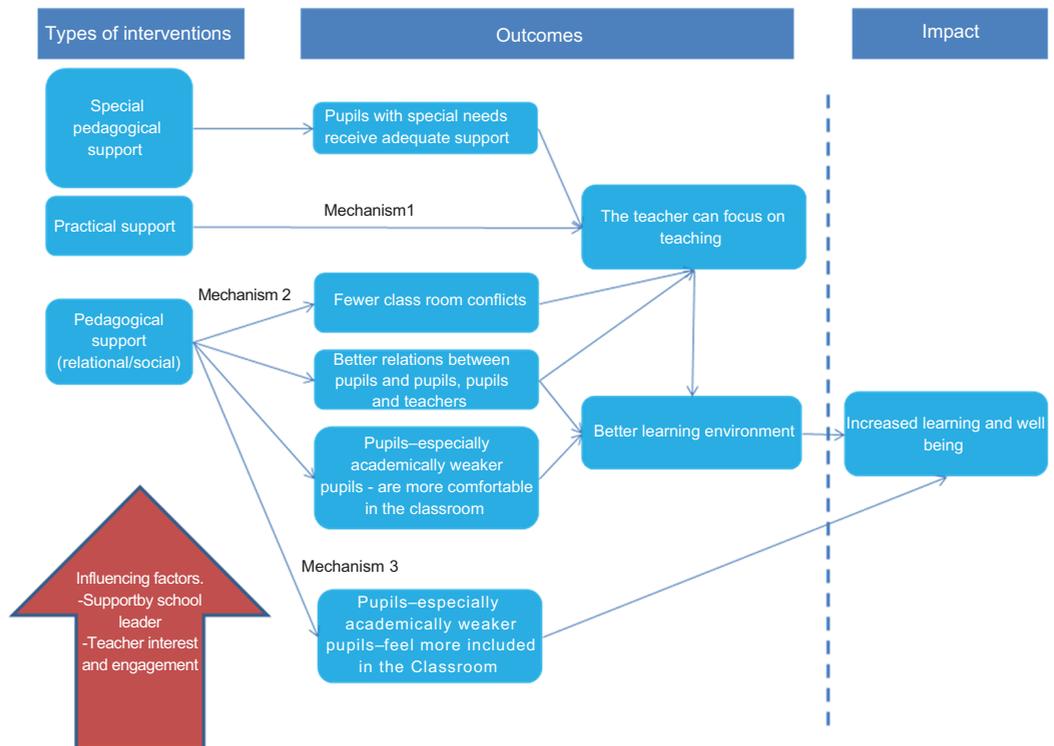


Figure 1. The initial program theory for the teaching assistant program.⁴

We carried out the evaluation in two stages, which corresponded to the two successive implementation years of the interventions. The two-stage implementation of the interventions was ideal for the intended – but often difficult-to-realize – iterative process of CA (see Dybdal et al., 2010). The aim of the first phase of the evaluation was to work through the first four steps of the CA, namely: setting out the cause-and-effect issue to be addressed; developing the postulated theory of change and the risks to it; gathering the existing evidence on the theory of change; and assessing the contribution story and the challenges to it (Mayne, 2011). During the first phase of the evaluation, we developed and refined the theory of change for each of the intervention types on the basis of the existing literature and a preliminary round of data collection and analysis.

In the second phase (the revision and strengthening phase), we paid particular attention to seeking out additional evidence and strengthening and finalizing the contribution story. This corresponds to Mayne's steps five through seven (2011). The data collection in the second stage of the evaluation involved:

- a survey of all school principals, relevant teachers and teaching assistants at the participating schools;
- a time series study on the impact of the teaching assistant interventions among those students who had maintained sustained contact with a teaching assistant. 'Impact' was defined as positive relation with teaching assistants, plus the pupils' classroom comfort and learning. The time series study involved three consecutive measurements over the course of the school year;
- case studies at six selected schools, involving individual and group interviews with school principals, teachers and teaching assistants, plus classroom observation.

The two-stage implementation process combined with the application of the different types of intervention at multiple sites granted us a unique opportunity to identify and examine possible influencing factors and alternative explanations. However, the strategies and experiences presented in relation to the REF are also applicable to many one-off, single-site evaluations.

As mentioned earlier, the application and use of the REF may differ from evaluation to evaluation. In our evaluation of the teaching assistant program we employed the REF during the final analysis phase during the second round of teaching assistant interventions. As part of evaluating the first round of teaching assistant interventions, we had identified a total of 12 salient potential explanations on the basis of data collection and analysis as well as existing literature. In addition to these explanations, we also identified five of the most relevant influencing factors. We first listed these in the framework, presenting the explanations first, followed by the influencing factors. Subsequently, we defined the type and level of each of the explanations and influencing factors. Two members of the evaluation team formulated predictors for each of the explanations and influencing factors. The entire evaluation team then commenced the analysis phase, using the specified predictors as well as the five 'degree of influence' criteria as a guiding structure for our analysis of qualitative and quantitative data. As we advanced through the analysis, we gradually compiled our evidence or counterevidence for each of the explanations and factors in the relevant spaces in the framework. Finally, we collectively discussed our findings on each of the explanations and influencing factors, summarized the implications of our findings for the contribution story and, as part of the evaluation, integrated our conclusions into the other analyses.

The results of our application of the REF are presented in the table below. For the purpose of this article we have selected three alternative explanations and two influencing factors to illustrate our application of the REF.

Table 2. (Continued)

1. Description	2. Type	3. Level	4. Identifiers	5. Degree of influence	6. Implication			
Alternative explanation/ influencing factor	Mechanism		Certainty	Robustness	Prevalence	Theoretical grounding		
A. <i>Support from the school leader is an influencing factor.</i>	Sustained support and attention to the program from school leaders will support its implementation and outcome.	Implementation	Interventions with sustained support and attention from the local school leader are related to positive results in relation to the TA experience of the program and / or the pupils' learning.	Moderate. Interventions with sustained support and attention from school leaders generate positive results for the TA process. The TA experience of the planning process. The TA experience of the implementation of the program and / or the pupils' learning is unaffected.	Low. The explanation was not identified as a contributor across data sources and methods.	Moderate. The explanation only accounts for an improvement in the early stages of implementation.	Low. There are no relevant studies.	The support of primarily influences the planning stage of the intervention. The influence does not translate into the actual implementation of the intervention in the class room.
B. <i>The interest of the teacher and their ability to collaborate with the TA is an influencing factor.</i>	Active collaboration between the teacher and TA can influence the implementation and outcome of the program.	Interpersonal	TA interventions with high collaboration between the teacher and the TA indicate positive results regarding the TA experience of the program.	Moderate. Interventions with high collaboration between the teacher and the TA indicate positive results regarding the TA experience of the program. TA experience of the implementation.	Moderate. Teachers and teaching assistants identify the explanation as a contributor.	Low. The explanation only accounts for an improvement in the implementation of the intervention.	Low. There are no relevant studies.	The support of the program is enhanced by active collaboration between the teacher and the TA, especially in interventions focusing on relationship and pedagogical support. These types of intervention typically require more collaboration.

The remainder of the section describes our experiences of working through each component of the REF.

Description of explanation and factor. The first step in applying the REF is lining up the potential explanations and factors. But what possible explanations exist for the observed outcomes? What contextual factors are likely to moderate or influence the underlying mechanisms? What might inform or help the evaluator in making this list? CA methodology papers suggest a wide range of informational sources, including stakeholders (Mayne, 2011), prior evaluations of the same or a similar program (Dybdal et al., 2010), prior evaluations and studies conducted on the same target population (Dybdal et al., 2010), prior evaluations and studies concerning the same outcome (Dybdal et al., 2010), prior evaluations concerning the same mechanism (Pawson, 2002),⁵ and/or implementation research (Dybdal et al., 2010; Mayne, 2011).⁶ Moreover, we advise triangulating these sources in order to reduce the possibility that carelessness or bias will result in overlooking salient explanations and factors.

The aforementioned sources enable us to identify factors and possible explanations, which we can then compile into a gross list of hypotheses for further investigation. However, in reality, many evaluations operate under contractual, temporal and economic constraints, which imply that the potential focal areas must be assessed carefully and must also correspond to the purpose and intended usage of the findings. Our experience suggests consolidating a highly selective list of the most salient factors and explanations before advancing to the next steps in the REF. This list may be prioritized using three different criteria: (i) their fit with the purpose of the evaluation; (ii) their importance to the overall contribution story; and (iii) their utility and importance to the stakeholders.

As emphasized above, the identification of relevant explanations and factors should take place in the early stage of the evaluation, before advancing the data collection and analysis. However, one challenge we encountered in our application of the REF was the emergence of additional explanations and identifiers, and we struggled with how we should consider these emerging explanations. Ultimately, we found that the emerging explanations often modified the alternative explanations and in effect enriched the contribution story.

We also faced the challenge of determining the number of influencing factors and alternative explanations to include in the analysis. This issue is not confined to the application of the REF, and is in a way like asking 'How long is a long piece of string?' There is no cut-and-dried answer to this question. However, our experience in relation to the teaching assistant interventions was that we simply reached a point of saturation in our identification of relevant factors and explanations, where no new alternative explanations emerged.

Type of explanation and factor. The second step in the REF was to identify the different types of potential explanations and factors. One challenge in making this identification concerns the delineation between the primary explanation(s) and the competing alternative explanations. The problem appeared in relation to the teaching assistant interventions because the interventions were developed on a very limited knowledge base (i.e. there are few preexisting explanations for why this type of teaching assistant intervention works). Moreover, the implementation of the interventions involved a combination of roles for the teaching assistants. Accordingly, the potential for a well-grounded delineation between primary and alternative explanations appeared somewhat weak.

This challenge, often overlooked in theory-based evaluation, becomes very apparent when using the REF. The REF pushes the evaluator to confront this challenge and carefully consider the

extent to which a primary explanation can be identified on a firm basis. The usability of the REF, however, is not contingent on the identification of a primary explanation at the onset of the evaluation. In conducting our analysis of the teaching assistants, we simply treated all relevant explanations as competing mechanisms, finding no firm basis for identifying a real primary explanation.

Level. Assigning the level of each of the influencing factors proved relatively straightforward and informative in terms of judging their coverage across the various levels. The application of level appeared less informative in relation to proposed mechanisms, as the mechanisms primarily functioned at the individual level. In future applications of the REF, it would be interesting to examine how the REF fares with different levels of mechanisms, such as Astbury and Leeuw's typology of situational, action-formation and transformational mechanisms (2010: 371).

Identifiers. We were able to develop relevant and measurable identifiers – predicted data patterns – for all the alternative explanations. However, the degree of specificity of our identifiers varied across the explanations. While some consisted of relative generic predictions (e.g. the well-being scores being higher among children in teaching assistant interventions focusing on conflict resolution), other predictions involved more detailed configurations (e.g. expected relationships between well-being scores and contextual factors, settings or target groups). In general, we found that the greater the level of specificity (or uniqueness) of the predictions, the stronger the foundation for judging the certainty of the match between predicted and observed data patterns. Informed by Trochim (1985) and his work on pattern matching, we suggest raising the specificity of data predictions by:

1. formulating differentiated patterns of change for different target (sub)groups (e.g. at-risk students versus regular students), different implementation settings (e.g. urban versus rural schools) or across different measures (e.g. well-being scores versus learning gain scores);
2. specifying relationships among mechanisms or between mechanisms and influencing factors;
3. specifying the predicted degree of influence in numerical predictions by use of point-estimates, cut-off scores or intervals (if at all possible); and
4. using existing theory to articulate more complex and finely varied mechanism-outcome patterns (to the extent possible).

The potential to employ one or more of these pattern-matching strategies is enhanced in evaluations characterized by planned variation, including multiple program implementations, multiple data collection methods, multiple target groups, and so on. All of these planned variations can be used to specify more detailed patterns in the data and in effect support more severe testing of the explanations.

Degree of influence. Central to the REF is gauging the relative influence of the selected explanations and factors across the five attributes of certainty, robustness, range, prevalence and theoretical grounding. Working with these five attributes proved both feasible and worthwhile, suggesting that each of the attributes deserves to remain in the framework.

However, the certainty of the findings – the extent to which the observed pattern matches the one predicted – was in some instances difficult to judge. This is due especially to the degree of specificity of some of our propositions, and the degree of codification and fidelity of implementation of the teaching assistant interventions. We found that the interventions had developed over time in certain local implementations, slowly approaching the boundary that separates the practical

and relational types of intervention (see the typology above). This in turn complicated our ability to relate data patterns emerging from these projects to some of our predictions. In effect, for some of the explanations the degree of certainty became hard to gauge.

In our experience with the REF, the solution to this challenge rests primarily on the ability of the evaluator to specify detailed data patterns. By raising the level of specificity of the predicted data pattern, the evaluator gains firmer ground from which to base his/her judgment (see suggestions above). While increasing the level of detail in the predictions involves more fine-grained analysis, the resultant CA is also enriched. As such, the additional level of analysis is, we believe, well worth the effort.

In the analysis phase of the evaluation, the robustness, range and prevalence criteria collectively represented a useful and relevant framework for dealing with the selected influencing explanations and factors. However, the robustness and prevalence criteria surpass the range criterion in usability and importance, as the latter was less applicable in relation to several of our alternative explanations. Systematically working through these attributes resulted in a transparent chain of evidence for each of the proposed explanations and factors.

Implication. One challenge with the REF is how to summarize the judgments regarding each of the five criteria, and how to weight these judgments across the different attributes. Our approach to this challenge has been a pragmatic one. We decided to use a simple scale (low, moderate, high) supported by statements concerning each of the four criteria. We then used this scale, plus the statements, in summarizing our judgment of the relative importance of the alternative explanations and influencing factors; often, we added clarifying or modifying descriptions to supplement our judgment. The use of a more elaborate scoring system or coding procedure could potentially make sense in other evaluations. The unbundling of rival explanations is not scripted, and there is no magic wand for doing so.

Conclusion

In this article we have examined two salient aspects of the methodology of CA, namely how to account for both influencing factors and alternative explanations. We have argued that in its current form, CA needs further methodological and practical elaboration. First, we clarified key terms concerning mechanisms, influencing factors and alternative explanations. Subsequently, we proposed a framework, the REF, for facilitating a systematic and transparent approach toward identifying and prioritizing the most important influencing factors and alternative explanations that need to be accounted for in CA.

Although some practical challenges emerged in our application of the REF, we also identified several strengths.

The use of the REF allowed us to establish a chain of evidence that spanned the early identification of alternative explanations to the final articulation of the contribution story. Accordingly, the component steps underpinning our examination and conclusions regarding the influencing factors and alternative explanations were transparent to the commissioner of the evaluation and the stakeholders involved.

By identifying the key strands of explanation to be pursued, the REF also provided a focusing structure for our data collection and analyses. The predictions made for each of the explanations and factors served as valuable leads for the data collection. In the analysis stage of the evaluation, the REF became the focal point at which the different strands of the contribution story were collectively summarized.

Finally, the REF informed our predictions about the future trialability of the teaching assistant interventions by suggesting how the interventions might work in other settings and contexts. We were able to use our judgments regarding robustness, range and prevalence for each of the alternative explanations and influencing factors in order to identify which components of the interventions must be faithfully implemented, and which ones allowed for local modification in future implementations.⁷

We have previously argued for the importance of the availability of a more operational framework for dealing with influencing factors and alternative explanations. In our minds, the absence of such a framework implies relying on less robust practices, which will potentially result in a credibility deficit for CA. The intended outcome of this article is that the REF will be employed and refined in other CA-based evaluations.

We encourage the pursuit of the following questions in future applications. Will the REF demonstrate relative advantages in terms of enhancing the credibility of contribution stories (over and above doing something else, or doing nothing at all)? What, if any, are the client benefits of using the REF? What, if any, unintended outcomes result from the practical application of the REF? How will the REF fare as a framework for the analysis of primary (or subsets of primary) explanations in more complex evaluative settings? Is the REF appropriate in handling different levels of mechanisms? Is the REF too taxing to employ in more complex evaluative settings?

Admittedly, the REF is in its early stages of development. Our modest hope is that our presentation will stimulate further application and refinement of the framework.

Acknowledgements

The authors would like to thank the other contributors to this volume, the anonymous reviewers, and their colleagues at Rambøll for their constructive comments. The evaluation of the teaching assistant interventions was funded by the Danish Agency for Quality Assurance and Evaluation.

Funding

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

Notes

1. One might object to this challenge on the grounds that accounting for the influence of competing factors and explanations only pertains to evaluation studies that aim to determine the net effect of a program or policy (i.e. proving that a set of desired outcomes can be precisely attributed to a given intervention). It might accordingly be argued that the purpose of CA is fundamentally different in scope, in that it merely seeks to address the contribution of the program or policy being evaluated (i.e. it makes the best possible case that a program has contributed to a set of desired outcomes). As such, there is no need to account for the specific influence of competing factors and explanations in the context of CA. We beg to differ. If the CA involves any type of claim regarding the relative contribution of an intervention, some kind of story regarding other salient factors and explanations is necessary. Simply identifying and listing these factors and explanations does not suffice.
2. This step is similar to Trochim's pattern matching approach that 'essentially consists of an attempt to determine the degree to which observations correspond to or "fit" a theory' (1985: 575).
3. The criteria are inspired by Van Evera's thoughts on good theories (1997).
4. The presented theory of change covers the major strands of mechanisms assumed to comprise the teaching assistant interventions, but it has been simplified for the purpose of the article.
5. It may be difficult to identify the mechanisms due to the different ways that search functions are designed in different databases.

6. Although significant similarities exist in what comprises influencing implementation factors, it is useful to distinguish between what is being *implemented* and consequently what is being *evaluated*: (i) a policy, (ii) a program, (iii) an idea/innovation, or (iv) a technology or skill (Coffman, 2010). These different implementation research approaches are informed by different disciplines, and theories and methodologies consequently differ. First, policy implementation research, informed by political science (Winter, 2006), focuses on the attributes of the various actors in the implementation process and how, for instance, low-level bureaucrats shape and reshape policies (Lipsky, 1980; Winter, 2003), and in turn how their implementation is influenced by political superiors (Keiser and Soss, 1998; Langbein, 2000), various organizational arrangements (Hill, 2006), administrative emphasis of policy goals (Ewalt and Jennings, 2004; Hill, 2006; Riccucci et al., 2004). Second, program implementation research, informed by such areas as the health, psychology, and education sciences, tends to focus on elements connected with fidelity to the program model (Durlak and Dupree, 2008; Dusenbury et al., 2003; Fixsen et al., 2005). Third, innovation implementation, informed, for instance, by communication, anthropology and sociology, focuses on the intended and unintended diffusion of innovations through communication and social systems (Rogers, 2003). Finally, technology implementation, informed by communications theory, business, and information science, tends to focus on conditions in the environment that enable the uptake of new technology in organizations (Grover, 1993).
7. As just one example, our REF-based analysis identified the specification of a clear role for the teaching assistant as well as teacher support as important influencing factors that augment the potential outcome of the teaching assistant. This information has been used to inform schools about how to best make use of teaching assistants (i.e. be specific in terms of their responsibilities and collaborate with teachers who support the idea of a teaching assistant in the class room).

References

- Astbury B and Leeuw F (2010) Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation* 31(3): 363–81.
- Coffman J (2010) Broadening the perspective on scale. *Evaluation Exchange* XV(1): 2–3.
- Durlak J and DuPre E (2008) Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology* 41: 327–50.
- Dusenbury L, Brannigan R, Falco M and Hansen W (2003) A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research* 18(2): 237–56.
- Dybdal L, Nielsen SB and Lemire S (2010) Contribution analysis applied: reflections on scope and methodology. *Canadian Journal of Program Evaluation* 25(2): 29–57.
- Ewalt JAG and Jennings ET (2004) Administration, governance, and policy tools in welfare administration. *Public Administration Review* 64: 449–62.
- Fixsen D, Naoom SL, Blasé KA, Friedman RM and Wallace F (2005) *Implementation Research. A Synthesis of the Literature*. Tampa, FL: University of South Florida.
- Grover V (1993) An empirically derived model for the adoption of customer-based interorganizational systems. *Decision Sciences* 24(3): 603–40.
- Hedström P and Swedberg R (1996) Social mechanisms. *Acta Sociologica* 39: 281–308.
- Hedström P and Swedberg R (eds) (1998) *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
- Hill CJ (2006) Casework job design and client outcomes in welfare-to-work offices. *Journal of Public Administration Research and Theory* 16: 263–88.
- Keiser LR and Soss J (1998) With good cause: bureaucratic discretion and the politics of child support enforcement. *American Journal of Political Science* 42: 1133–56.
- Langbein LI (2000) Ownership, empowerment, and productivity: some empirical evidence on the causes and consequences of employee discretion. *Journal of Policy Analysis and Management* 19: 427–49.

- Lipsky M (1980) *Street-Level Bureaucrats: The Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.
- Mayne J (1999) Addressing attribution through contribution analysis: using performance measures sensibly. Discussion paper. Office of the Auditor General of Canada. URL: <http://dsp-psd.pwgsc.gc.ca/Collection/FA3-31-1999E.pdf>.
- 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 16*, Institutional Learning and Change (ILAC) Initiative, Rome, Italy. URL: http://www.inteval-group.org/IMG/pdf/ILAC_Contribution_Analysis-2.pdf.
- Mayne J (2009) Improving HRSDC evaluations of G&C Program: using theory-based approaches. Unpublished manuscript.
- Mayne J (2011) Addressing cause and effect in simple and complex settings through contribution analysis. In: Schwartz R, Forss K and Marra M (eds) *Evaluating the Complex. Attribution, Contribution and Beyond*. New York: Transaction Publishers.
- Mayntz R (2004) Mechanisms in the analysis of social macro-phenomena. *Philosophy of the Social Sciences* 34: 237–59.
- Pawson R (2002) Evidence-based policy: the promise of ‘realist synthesis’. *Evaluation* 8(3): 340–58.
- Pawson R and Tilley N (1997) *Realistic Evaluation*. London: SAGE.
- Pawson R, Greenhalgh T and Harvey G (2004) Realist synthesis: an introduction. *ESRC Research Methods Programme*, University of Manchester RMP Methods Paper 2/2004, pp. 7–8. URL (consulted May 2011): <http://www.ccsr.ac.uk/methods/publications/documents/RMPmethods2.pdf>.
- Riccucci NM, Meyers MK, Lurie I and Han JS (2004) The implementation of welfare reform policy: the role of public managers in front-line practices. *Public Administrative Review* 64: 438–48.
- Rogers EM (2003) *Diffusion of Innovations*. New York: Free Press.
- Trochim WMK (1985) Pattern matching, validity, and conceptualization in program evaluation. *Evaluation Review* 9(5): 575–604.
- Van Evera S (1997) *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.
- Weiss CH (1997) Theory-based evaluation: past, present, and future. *New Directions for Evaluation* 76: 41–55.
- Winter S (2003) Political control, street-level bureaucrats and information asymmetry in regulatory and social policies. Paper presented at the Annual Meeting of the Association for Policy Analysis and Management held in Washington, DC, 6–8 November. Copenhagen, Denmark: The Danish National Centre for Social Research.
- Winter S (2006) Implementation. In: Peters BG and Pierre J (eds) *Handbook of Public Policy*. Thousand Oaks, CA: SAGE, 151–66.
- Yin RK (2003) *Case Study Research. Design and Methods*. Thousand Oaks, CA: SAGE.

Sebastian T. Lemire is a senior consultant at Rambøll Management Consulting in Denmark. He holds an MSc in social science methodology from the London School of Economics. He has carried out a number of evaluations, mainly in the educational sector in Denmark and the USA. His current research interest lies in methodologies that address the attribution question, including theory-based evaluation.

Steffen Bohni Nielsen is Director of Evaluative Knowledge at Rambøll Management Consulting. His main research interests are evaluation capacity building, theory-based evaluation and results-based management.

Line Dybdal is a manager and team leader at Rambøll Management Consulting. She has carried out numerous evaluations in the field of at-risk families. Her research interests include theory-based evaluation and evidence-based policy and practice.