

CHAPTER 10

Evaluation as a Tool for Leading Change

PASI-HEIKKI RANNISTO & ANNA SALORANTA

Takeaways for Leading Change

In this chapter, we acquaint the reader with evaluation, touching shortly on its functions and some of the theories informing it, as well as the role it plays in the changing public sector. We introduce practical ways of using evaluations as tools for managing change. Finally, we argue these are tools we already often have at our disposal but do not utilise fully. Leaders may not recognise the opportunities and tools an evaluation provides them with in terms of leading change. We hope to show how embracing a wider culture of evaluation will build a foundation of systematic and shared information helping leaders create and discuss change across sectors.

Effecting change in complex and challenging contexts requires tools in addition to leadership skills. Evaluation is one such tool; it offers a functional point of view of leadership and the practical management of change by describing complex environments, relationships, and problems from the perspectives of all stakeholders. Evaluation does not only show whether the progress achieved is in line with project objectives. It can also be used to educate the organisation or network about an ongoing process of change. Such a process provides an opportunity to define and communicate leadership visions and goals as they pertain to individuals, processes, and organisational structures.

Change, evaluation, leadership, and management share many of the same functions and operation models, most notably the setting of objectives and the defining of practical outcomes and measurements. Several studies have shown that leading change efficiently requires evaluation. In practice, few organisations do this. Consequently, they never find out why the desired change has manifested as something undesirable or why their outcomes do not match their designs (Doyle, Claydon, & Buchanan, 2000; Hayes, 2002; Kirkpatrick, 1985; Owen & Rogers, 1999; Skinner, 2004). Both change projects and evaluations are central themes in results-based management. Their key factors are achieving development targets, clarifying the roles and responsibilities of actors, increasing transparency and accountability in functions, and using good-quality data to improve decision-making (Rist, Boily, & Martin, 2011).

Our approach to evaluation and its use in leading change is broad, requiring us to examine our subject from many different angles. First, evaluation can be seen as a means of measuring performance. The objective can vary from the success of an individual development or change project, operational work, organisational restructuring, and implementation of organisational strategy, to the success of a policy on a national or local level (Figure 1). Second, evaluation is needed when leading the process of change itself.

Evaluation can be done for a number of different purposes. In education, for instance, evaluating students is a key process for teachers. Based on the information gained, a teacher might decide to teach in a different way or give extra support and encouragement in certain areas. In social work



FIGURE 1. Different foci of evaluation

or healthcare, the effects of interventions and treatments in the lives and behaviours of individuals is under constant scrutiny. In this context, the process shows the connections between the chosen intervention, the individual's behaviour, and the outcome, all in relation to the original objective. The change, development, and healing of an individual is followed closely during care, as are its effects on his or her life. In these instances, evaluation is a natural part of a service process, like a visit to your healthcare provider or an exam as part of a learning process.

Evaluation can also be directed towards the outcome of a larger government reform or policy. This can be done from a number of different perspectives, such as that of the citizen, politics, or government (Airaksinen, Haveri, & Vallo, 2005). The different roles of citizens can be the object of interest. They are voters, decision makers, service users, and taxpayers. The political viewpoint would entail examining the political administrative systems as well as the division of labour between different actors. From a governmental point of view, one would focus on the administrative culture and the work of civil servants as well as on the government as a whole. Another point of interest might be the objectives stated in legislation and how these have been implemented by the government (Temmes, 1994).

Reforms are often evaluated in different ways throughout the stages of their implementation. From the very beginning, the focus would be on the design, objectives, and available resources. At the point of completion and even past that, the focus would be on how the original objectives were met and whether the change had the desired effects. A reform can also be

evaluated during its implementation, the object then being the process of change itself and the actions taken to achieve it. It has been suggested that all government reforms should undergo evaluations from their planning stage onwards (Skinner, 2004; Temmes, 1994).

A Short History of Evaluation in the Public Sector

Three distinct stages can be identified in the development of evaluation in the public sector, all of them linked to a change in welfare state policy making (Wollmann, 2003). From the 1960s onwards, there has been increasing interest in public policy-making and its implementation. In both the academic and the public spheres, it has become clear that formulating policies and funding programmes are simply not enough to effect societal change. There has to be accountability and some idea of how programmes work. Research and evaluations focused on how policies are taken from plan to reality. Many factors have been identified as having an impact on the process of implementation, not least of which was the effect of individual actors themselves (for an introduction to policy implementation, see, for instance, Pressman & Wildavsky [1984] or Sabatier [1991]).

However, research on implementation alone did not sufficiently inform policy-makers of the efficacy of their policies. Examining outcomes as well as implementation was necessary for sufficient feedback for policy reformulation. The goal was to improve not only societal discrepancies but also the policies addressing them. This was the beginning of the second stage of development in the 1970s and 1980s (Rist et al., 2011; Wollmann, 2003).

Simultaneously, New Public Management (NPM) strongly affected the execution of public policies. They were to be brought under the same kind of efficiency and output requirements as the private sector and industry. NPM strove to bring about a more efficient and economic thinking to public sector operations and management (Pollitt, 1995). This meant that both the policy outcomes and formulation were examined, changing the way evaluations were conducted. Merely measuring output does not satisfy the

question “Is this policy efficient?” In the 2000s, evaluation has developed alongside the metamorphosis of NPM into New Public Governance (NPG), bringing about the third stage of development.

During this period, policy-makers have sought to adapt to growing changes in increasingly complex environments. Instead of policies or projects being implemented from the top down, the emphasis has been on shared responsibility and shared decision-making between partners. Controlled public-private partnerships are evolving into networks. Top-down implementation has given way to co-creation. Partnerships are no longer seen as being formed solely between organisations but now also include citizens, either as active individuals and customers or as non-profit organisations and interest groups (Christensen & Laegreid, 2017). Not only has the public sector changed in the Western welfare state, especially in the Nordic countries, but democracy itself is also finding new ways of expression. The rate of change is challenging and it changes the way we have to think about and conduct evaluations. To produce the information necessary for effective leadership in this complex context, evaluations need to factor in a wider range of stakeholders, operational cultures and networks. The range of different identifiable mechanisms possibly effecting the outcomes becomes much wider.

The outcomes of varying combinations of contexts and mechanisms offer evaluation researchers the building blocks for testing different theories. There is no one true evaluation theory but several coexisting ones. Evaluators are responsible for clarifying their theoretical approach and the methods they have chosen in their own work. One of the theories commonly used for complex evaluations is, the realistic evaluation framework. It is popular today because it accommodates a variety of methods, thus making it easier to measure more complicated changes. The “fathers” of realistic evaluation are Pawson and Tilley (1997). They developed the approach around a basic theory-driven concept of Mechanism + Context = Outcome. This means that the building of the evaluation framework starts with a theory of causal explanation. The regularities of patterns in social activities form the mechanism, and all human activities happen in a specific context.

Top-down
implementation
has given way
to co-creation.

The realistic evaluation cycle is built around testing for variations in variables in an attempt to expand the underlying theory. The testing itself, according to Pawson and Tilley (1997), is done with the method best suited for answering the research question.

Realistic evaluations are done in cycles, each one adjusted by the knowledge acquired in the previous cycle. When evaluating a social intervention, it senses to focus first on statistical analysis to get the big picture. Then, in consecutive cycles, focus more deeply on more interesting aspects by using qualitative methods (e.g., interviews) in a specific community/part of the population.

The specific value of realistic evaluations is that the design is adaptable to the course and needs of the project. There are various opportunities to educate and engage the actors and stakeholders during the different cycles of the evaluation. Ultimately, it is the responsibility of evaluators to ask and answer questions that will most benefit the project at that particular time. These could be questions about what the preconditions for a successful implementation are, questions about how a project is being implemented and where it is likely to lead, and questions about what kinds of outcomes a certain project has produced and whether these are in line with the objectives.

Evaluation as a Tool for Creating and Leading Change

The theories of change management and realistic evaluation share some commonalities, not least of which is the fact that each evaluation process or change is unique. Generalisations only work when operating with a broad analytical frame of reference that is comprehensible to all involved. Without a well-thought-out design, neither change management nor evaluation will produce sufficient information to determine success or enable learning (Kuipers & Richardson, 1999; Skinner, 2004).

Evaluations and processes of change are affected by the context and conditions. Both attempt to create a systematic method of implementation

and approach to the objective. In the former, the attempt is to conduct the evaluation process itself. In the latter, it is to effect change (Preskill & Torres, 1999; Russ-Eft & Preskill, 2009). Context is an important factor when evaluating different changes and interventions or when trying to generalise and learn about causal pathways (Blamey & Mackenzie, 2007).

Both change and evaluation approaches share a disinclination to use qualitative methods as the only data collecting strategy for the evaluation of complex interventions or changes. One reason for this could be that even though the participant's experience and view, as measured by qualitative methods, produce an estimate of the effects of conditions particular to a project, the results will not be relatable to any other project. The results are relative. For this reason, using qualitative methods might leave the question of what works in the greater context of things unanswered (Blamey & Mackenzie, 2007). This is not to propose that qualitative methods are obsolete, but rather that being aware of the possibilities and restrictions of the methods one intends to apply is central. The material or data for evaluation can be obtained from many sources. A secure approach is to collect data from numerous sources, both quantitative and qualitative. This enables data triangulation in order to ensure the reliability of the results.

In the management of change, the final result is the focus of all attention, whereas this is only part of what falls under an evaluation, where the processes, different conditions, steering mechanisms, and management are all of interest and examined with the eyes of an outsider. The objectives of a project or reform direct people's actions as well as the evaluation of the progress. Operations and processes are measured and can be realigned if they are not in line with the set objective. Furthermore, evaluation will examine the change as a whole during the entire process, while management of change looks forward to the end result. In the management of change, decisions are made for an assumed future reality, when evaluation can operate more or less in reality (Blamey & Mackenzie, 2007; Skinner, 2004).

By striving for economy, effectiveness, and practical solutions, processes of change and evaluation aim at finding new ways to function by looking for unsuitable ways of action and changing them. Measures central to reaching the set objectives can be found naturally in these processes (Blamey & Mackenzie, 2007; Pawson & Tilley, 1997). Pawson

and Tilley (2001) emphasise the importance of developing a culture of evaluation and the perseverance required to do so. They add that the traits of the evaluation factors also work well in managing organisational change. When problematic themes arise during an evaluation, their origins may, after further inspection, already be visible in earlier evaluations. Being able to rely on several consecutive evaluations or evaluation cycles helps to identify the root causes of a problem, thus enabling leadership and management to tackle the disease instead of the symptom. This plural nature is also true for the management of change. Each process of change already bears within it the seed for the following one. Some things will change in the intended direction, but something else will fail to change. Development through evaluation is cyclical, not linear (Bouckaert & Halligan, 2007).

During the evaluation, small questions are worth asking, even to test big theories. In evaluation, as in change, it is worthwhile to proceed with small steps even if the overall picture is big. It is important to use as many different methods as possible. This way, a clearer, more versatile picture can be obtained from the actual outcome. It is important to notice which mechanisms are relevant to producing optimum outcomes by context. Policies affect several practical mechanisms, changing and directing them. It is possible for different environments and conditions to produce similar results. Meta-analytic inquiries may help to make common policy mechanisms more visible. Asking the “why” questions and trying to answer how the conditions have affected the outcomes is important (Pawson & Tilley, 2001).

As leaders and evaluators, we need to be careful never to assume to know what works. Instead, we must persistently try to prove it. People act differently, conditions change, and new points of view arise. These can all have an effect on what works in a given situation. In both evaluation and managing change, the focus should be on finding out what genuinely works and why (Pawson & Tilley, 2001).

Leadership makes up its own level of evaluative inquiry. Today, many national policies have goals such as the change of organisational culture. This means changing the way different professions or departments communicate, form networks, and acting around shared processes

for the benefit of shared goals and/or clients. These are truly complex environments. From a leadership point of view, creating a shared culture of operation in a multiorganisational context poses new demands, as the actors involved are not under shared management. Instead, it has a number of leaders, all representing their own professions. Evaluations can provide a tool for shared objective setting and the defining of goals, as well as a natural platform for discussion and education. Different studies have pointed out the necessity of feedback in gaining acceptance of, and commitment to, change objectives. Evaluation done from the beginning of the change process can provide staff with the opportunity to tell the management and leadership about its own experience; it also offers development proposals for the implementation of change (Carnall, 1990; Kirkpatrick, 1985; Preskill & Torres, 1999).

Evaluation in Practice

The reason that evaluations have become so common may lie in the direction in which the public sector has evolved, requiring projects to quickly and efficiently source information and gauge their starting point. The choice of what to measure to determine the success of a project is not trivial or simple. Between deciding what to measure (quality, customer value, output, or input) and what point of view to choose (customer, employee, service provider, or service commissioner), it can be hard to show indisputably the effects of any process.

Evaluation is always connected to some purpose or objective. Without this, it is often difficult to see the difference between a success and a failure in complex situations and environments. Successes and failures can be difficult to relay without the information provided by an evaluation process (Patton, 1987). At best, evaluation will provide new information, points of view, and solutions that can be used in the next phase of a reform or organisational change (Chelimsky & Shadish, 1997). At its core, it is a question of making value judgements and assessments to reach a conclusion in terms of success. This is done through a regulated process wherein a phenomenon is examined through a set of evaluation criteria (Vedung, 1997).

The data used are often collected from numerous sources in both qualitative and quantitative forms. This helps to ensure the reliability of the results. It is also important to handle data to ensure the respondents' privacy and the transparency of data collection and use. Even though the evaluator strives for neutrality, in reality, he/she is often comparable to an active change agent. It is also worth noting that the more the evaluator participates, the more the evaluation project starts to become a driver for change (Kuipers & Richardson, 1999).

The processes and outcomes of evaluations can be seen as interactive constructs of meaning defined by evaluators and evaluatees alike. Underlying this are the intentions, norms, values, and actions of both. Studies have found that the central characteristics of evaluation practices are as follows (Chen, 1990; Kuipers & Richardson, 1999; Patton, 1987; Scriven, 1991):

- When evaluating a process of change, the objectives of the evaluation will evolve along with the process. As a result, the organisation itself will also change.
- The aim is to generate integral knowledge and to communicate it to create a broader understanding of the object.
- Evaluation is messy.
- Evaluation results entail multiple processes including negotiations and communication.
- The object of an evaluation is specific to time and place, as it is constantly evolving.
- Evaluation is formative, and its purpose is to improve and make participants aware of its goals and developments.
- The outcome of an evaluation process ought to be relevant, acceptable, and understandable.
- Starting points of evaluation include implicit or explicit theoretical notions of goals, effects, processes, and so on.
- A mix of research strategies, methods, and techniques (both qualitative and quantitative) can be used.

Policies and strategies are often evaluated *ex ante* or *post hoc*. An *ex ante* evaluation is carried out to make some sort of prediction about how a policy change or a new policy will have the desired effect, as well as

to identify possible unwanted effects. A post hoc policy evaluation will speak to the original goals versus outcomes as well as the implementation itself – for instance, in terms of compliance with national and international regulations. The task is often put to tender by a governmental body under whose jurisdiction the policy falls. The way in which the tender is formulated in terms of goals, objectives, and time frame will invariably affect the focus and methods of the evaluation.

A thorough evaluation at the policy level requires starting from the planning phase and the first stages of the implementation of the reform. This usually means an internal process is begun alongside the planning phase of the project to ensure not only a reliable evaluation of the first stage, but also that all relevant material is available for upcoming evaluations. Ideally, in a post hoc evaluation, the change and its effects will be visible in all the different contexts in which it played out. In complex environments, this means a variety of stakeholders, contexts, and consequently, material. It makes sense to have a number of evaluations before and during a policy reform, and then to gather all the results at some point after its implementation. It is worth noting that different phenomena of the change cannot be constructed after the fact if the material has not been collected in real time and in context (Skinner, 2004).

Evaluation at the operative level speaks to the local and regional governments and their ability to affect change. It can be done before, during, or after implementation, both internally and externally. The previous level was concerned with creating the necessary conditions and incentives to bring about change. The operative level is the context in which change is created. Here, policies turn into flesh and blood. The relationships and workings between governments, counties, joint authorities of various kinds, and local politics and municipalities form abstract policies into concrete plans.

It is at the level of services that a public sector reform can really be measured. Here lies the basis for answering questions about costs, efficiency, and distribution of services. The organisation or production

The operative level is the context in which change is created. Here, policies turn into flesh and blood.

of the services, or both aspects, are most often the point of interest. It is easy for managers and leaders whose operations are being evaluated to feel as though they are under scrutiny. After all, it is their responsibility to implement a national or regional policy, and their performance data are being used to show whether this has happened. The evaluation process itself can encounter some resistance within an organisation or network; this is most often because individuals misguidedly feel as though they themselves are under scrutiny. It is critical that evaluators maintain a good and ethical relationship with the evaluatees throughout the process, and make the objectives and goals of the evaluation clear to minimise discomfort and resistance.

When evaluating operative organisations (e.g., municipal organisations and joint authorities) undergoing change, it is important to have access to data recording decisions and administrative solutions. These create the practical conditions for carrying out and managing change. The quick decisions and reactions common in projects are often left undocumented. This poses a challenge for evaluation, as it is often difficult to verify these later on. From the research point of view, this leaves the interpretation of the true reasons for the achieved outcome open to mistakes.

At the level of public services, evaluations are often done before, after, and during a reform and by both internal and external parties. This is not only because performance measurement is an integral part of both NPM and NPG, but also because it is important to gain a comprehensive picture of the change and its effects. Different evaluators seek to answer their own questions, gathering a diversity of materials that help to strengthen the overall evaluation.

Critique of Social and Healthcare Service Evaluation

There is some criticism among social and healthcare workers and researchers of social work against the idea of measuring effectiveness. NPM is thought to have brought about evaluation models that are too focused

on market values and might eventually serve to weaken the foundations of the welfare state (Brodkin, 2008; Mänttari-van der Kuip, 2013; Palola & Karjalainen, 2011; Rajavaara, 2007). This unwillingness to be measured can also be construed as an attempt to protect the profession and field of research from outside influences and criticism. Decision-making in social and healthcare services is often based on authority rather than information. Authority, according to Gambrill (2001), can be used to control reality through rhetoric. This creates a chasm between the rhetoric and proven outcomes, the filling of which would require systematic information gathering and measurement, as well as the evaluation of outcomes and effectiveness (Gambrill, 2001; Paasio, 2003).

Despite these criticisms and doubts, it is clear that systematic evaluations of public services will continue. Digitalisation and access to mass data provide quick and cost-efficient opportunities to estimate and evaluate the productivity and effectiveness of any form of decision-making. Alongside the professions, organisational management and politicians also have stakes in the provision and commissioning of services. By choosing opposition instead of participation, the professions run the risk of other arguably less informed stakeholders shaping the future of performance measurement in public services.

Discussion

The importance of evaluation as a tool for leading change is often underestimated. In this chapter, we have aimed to show some of its different uses in leadership questions. Taking the public sector as our starting point, we argued that the information provided by the almost ubiquitous evaluations in the sector can be harnessed to provide leadership with tools for anything from performance measurement to communicating visions and goals and creating learning in the organisation. To achieve this, timing is crucial.

Communicating visions and goals can be made a natural part of the objective setting and defining of evaluation criteria. Defining evaluation criteria is something the evaluators cannot do alone and will need the

input of the organisation for. Taking the question to different levels of the organisation will not only help to open up and clarify the criteria but will also serve as opportunities for leadership to communicate personally the goals of the change process. Meeting in person is a two-way process. The messages and information exchanged flow both ways. This means the leadership are given a chance to communicate their visions and goals and employees also get a chance to clarify what the envisioned changes entail for them.

Any process of change is likely to bring with it some amount of resistance. This may be expressed as active disagreement with the goals of the change or simply be caused by “reform exhaustion”—trying to implement too many changes at a time or in quick succession. To remedy this, employees need to be provided with the opportunity to discuss the issue with each other and with the management. That the superiors also have the opportunity to join the mutual discussion is essential. Systematic and transparent planning, measurement, and follow-up help to facilitate the lead-in of the change and of the evaluation processes (Doyle et al., 2000). Through an evaluation process, disgruntled factions in the organisation as well as its leadership and management are provided with information about the success of the process by a neutral third party. An evaluation will provide opportunities to educate the organisation about the different phases of the change process.

With the help of leadership, employees can become motivated to implement change and to learn new skills. Having an understanding of one’s environment and the support of one’s superiors helps in the learning of new skills, in taking on new information, and in the perception of causal relations within the organisation. All of this can bring about organisational learning. New and important fields of expertise and new practices may be discovered, leading to changes taking place in the work environment or in individual tasks (Blamey & Mackenzie, 2007; Doyle et al., 2000).

One of the most useful tools for leading change that evaluation offers is the opportunity to steer a process of change. This is not to say change processes are otherwise not steered, but merely to point out how well the evaluation process lends itself to be used for this purpose. Evaluation can be used to identify the need to change course. The initial signals that something needs to be fixed can be subtle and lost in the noise of

the other day-to-day information of the organisation. Having evaluation data, which has been analysed and approached differently compared with normal performance measurement data, may help in the early detection of problems. Information and signals are more easily and reliably identified by providing an information feedback cycle independent of and parallel to the organisation's own. A neutral third party providing this information often makes it more agreeable to sceptics of change, who are consequently more likely to engage in discussion about the process, providing a valuable perspective on the overall change discussion.

In conclusion, an evaluation alongside a change process can be thought of as a mirror held up to reflect an organisation over a given period. Leaders can use this to their advantage by picking up, acting on, and following up on the things the mirror reveals and perhaps comparing them to earlier images. How clear the picture is depends on how well the evaluation criteria and objectives have been set. This is an ever more demanding task as environments grow increasingly complex. In terms of leading employees, the mirror can serve two purposes. First, a leader needs to be visible. The magic mirror of evaluation will show just how, where, and when a leader is visible to employees. Second, the mirror provides leaders with opportunities to show and communicate the overall picture and its development back to the organisation. Naturally, evaluation is not the only tool that can do this, and some leaders do not need to rely on different phases of an evaluation to communicate efficiently with their organisations. However, we argue evaluation is often needlessly overlooked for this purpose in the search for more elaborate and impressive tools.

References

- Airaksinen, J., Haveri, A., & Vallo, M. (2005). *Seutuyhteistyön tilinpäätös* [Closing of the books for regional cooperation]. Tampere, Finland: Tampere University Press.
- Blamey, A., & Mackenzie, M. (2007). Theories of change and realistic evaluation: Peas in a pod or apples and oranges? *Evaluation*, 13(4), 439–455. <https://doi.org/10.1177/1356389007082129>

- Bouckaert, G., & Halligan, J. (2007). *Managing performance: International comparisons*. London, England: Routledge.
<https://doi.org/10.4324/9780203935958>
- Brodkin, E. Z. (2008). Accountability in street-level organizations. *International Journal of Public Administration*, 31(3), 317–336.
<https://doi.org/10.1080/01900690701590587>
- Carnall, C. A. (1990). *Managing change in organizations*. New York, NY: Prentice Hall.
- Chelimsky, E., & Shadish, W. R. (1997). *Evaluation for the 21st century: A handbook*. Thousand Oaks, CA: Sage.
- Chen, H. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Christensen, T., & Lægreid, P. (2017). *Transcending new public management: The transformation of public sector reforms*. Aldershot, England: Ashgate. <https://doi.org/10.4324/9781315235790>
- Doyle, M., Claydon, T., & Buchanan, D. (2000). Mixed results, lousy process: The management experience of organizational change. *British Journal of Management*, 11(Supplement 1), S59–S80.
<https://doi.org/10.1111/1467-8551.11.s1.6>
- Gambrill, E. (2001). Social work: An authority-based profession. *Research on Social Work Practice*, 11(2), 166–175.
<https://doi.org/10.1177/104973150101100203>
- Hayes, J. (2002). *The theory and practice of change management*. New York, NY: Palgrave.
- Kirkpatrick, D. L. (1985). *How to manage change effectively*. San Francisco, CA: Jossey-Bass.
- Kuipers, H., & Richardson, R. (1999). Active qualitative evaluation: Core elements and procedures. *Evaluation*, 5(1), 61–79.
<https://doi.org/10.1177/13563899922208823>
- Mänttari-van der Kuip, M. (2013). Julkinen sosiaalityö markkinoistumisen armoilla? [Public social work at the mercy of markets] *Yhteiskuntapolitiikka*, 78. Retrieved from <http://urn.fi/URN:NBN:fi-fe201306103881>
- Owen, J. M., & Rogers, P. J. (1999). *Program evaluation: Forms and approaches*. London, England: Sage.

- Paasio, P. (2003). *Vaikuttavuuden arvioinnin rakenne ja mahdollisuus sosiaalialalla* [The structures and possibilities of evaluating efficiency in the social sector]. Helsinki, Finland: Stakes.
- Palola, E., & Karjalainen, V. (2011). *Sosiaalipolitiikka: Hukassa vai uuden jäljillä?* [Social politics: is it lost or on to something new?]. Helsinki, Finland: Terveyden ja hyvinvoinnin laitos.
- Patton, M. Q. (1987). *How to use qualitative methods in evaluation*. Newbury Park, CA: Sage.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London, England: Sage.
- Pawson, R., & Tilley, N. (2001). Realistic evaluation bloodlines. *American Journal of Evaluation*, 22(3), 317–324.
[https://doi.org/10.1016/S1098-2140\(01\)00141-2](https://doi.org/10.1016/S1098-2140(01)00141-2)
- Pollitt, C. (1995). Justification by works or by faith? Evaluating the new public management. *Evaluation*, 1(2), 133–154.
<https://doi.org/10.1177/135638909500100202>
- Preskill, H. S., & Torres, R. T. (1999). *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage.
- Pressman, J. L., & Wildavsky, A. (1984). *Implementation: How great expectations in Washington are dashed in Oakland; or, why it's amazing that federal programs work at all, this being a saga of the economic development administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hopes*. Berkeley, CA: University of California Press.
- Rajavaara, M. (2007). *Vaikuttavuusyhdistykunta: Sosiaalisten olojen arvostelusta vaikutusten todentamiseen* [The outcome-oriented society. Styles of reasoning about the outcomes in social policy]. Helsinki, Finland: Kela.
- Rist, R. C., Boily, M., & Martin, F. (2011). *Influencing change: Building evaluation capacity to strengthen governance*. Herndon, VA: The World Bank. <https://doi.org/10.1596/978-0-8213-8403-9>
- Russ-Eft, D. F., & Preskill, H. (2009). *Evaluation in organizations: A systematic approach to enhancing learning, performance, and change*. New York, NY: Basic Books.

- Sabatier, P. A. (1991). Toward better theories of the policy process. *PS: Political Science and Politics*, 24(2), 147–156.
<https://doi.org/10.2307/419923>
- Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park, CA: Sage.
- Skinner, D. (2004). Evaluation and change management: Rhetoric and reality. *Human Resource Management Journal*, 14(3), 5–19.
<https://doi.org/10.1111/j.1748-8583.2004.tb00123.x>
- Temmes, M. (1994). *Hallinto puntarissa: Hallintouudistusten arvioinnin mahdollisuudet ja edellytykset* [Weighing up the government: the possibilities and prerequisites of government reform]. Helsinki, Finland: Ministry of Finance.
- Vedung, E. (1997). *Public policy and program evaluation*. New Brunswick, NJ: Transaction.
- Wollmann, H. (2003). *Evaluation in public-sector reform: Concepts and practice in international perspective*. Cheltenham, England: Edward Elgar.