20. Learning through evaluation: the missing link in governance of projects

Ole Jonny Klakegg and Gro Holst Volden

INTRODUCTION

Governance frameworks have existed from the beginning of project management, although the term was not used until much later. The current generation of governance frameworks for public sector investment projects have been developed over the last 20 years and have now become mature in several aspects. In particular, this yields the stage-gate structures with their control functions and decision-making. One aspect, though, has not matured to the same degree: the learning loop, based on *ex post* evaluation of completed projects, which is still not an integrated aspect of most governance frameworks.

Argyris and Schön (1978) developed the theory of action, in which they explained how agents are designers of action to achieve certain goals. By only focusing on the short-sighted degree of achievement, the theory explains how the agents are restricted to correcting their course to improve efficiency. This is called single loop learning. When the agents mature enough to consider the context in which the action is performed and its long-term effects, the theory explains how the agents also can consider what actions are effective and, thus, what actions are best fitted to given situations. This is called double loop learning.

Reflection is another important aspect of learning. As Greenwood (2003, p. 1049) pointed out, the purpose of reflective practice is to create more desirable outcomes. She refers to Schön (1983), who defined two components: reflection-in-action and reflection-on-action, where the latter corresponds to double loop learning. Double loop learning is necessary because the improvement may depend on structural as well as action redesign.

Building on 20 years of experience from the Norwegian quality assurance regime, this chapter illustrates the importance of *ex ante* appraisal, and *ex post* evaluations alike, in harvesting the benefits of having a working governance framework for public investment projects. Governing through all stages of development through front end and execution of projects is important but can only fix minor issues once a decision to finance is made. It is restricted to single loop learning ("doing it right," cf. Argyris and Schön, 1996) and adjusting the course.

In the long-term perspective, improvements resulting from double loop learning ("doing the right thing") are far more important. Did the projects achieve their goals for major user groups? Were they good value for money? Were they sustainable? If not, what needs to be changed? It is vital to pick up on changes in the project environment as well as shifts in policy and priorities, and even needs for structural changes in the governance framework itself. This chapter suggests necessary principles and qualities required in governance frameworks to achieve the desired learning effects.

THE LEARNING ORGANIZATION

Intuitively we accept that every organization has an ambition to improve and develop its capability pursuing some measure of success. What organizations aim to achieve is individual to each organization and situational, and it changes and develops over time. What remains constant is the need for further development. This implies the need for learning.

Learning in organizations has been a topic for authors and researchers over many decades. Örtenblad (2018, p. 151) points out the focus on learning in organizations comes from two directions. One can be described as "organized learning," building on pedagogy and educational studies, putting this into the context of management in an organizational setting. This puts the focus on management's ability to stimulate learning in the organization. The other direction can be described as "organizational learning," focusing on the organization where the learning takes place. This puts focus on the organization's need for development and learning, as well as organizational success factors to make learning possible.

Furthermore, Örtenblad (2018, pp. 151–152) discusses the definitions of the concepts of "learning organization" and "organizational learning." He claims that some kind of organizational aspect must be related to the learning. The organizational aspect may take the form of:

- the organization being a facilitator, supporter, and/or arranger of the learning going on in the organization, performed by individuals ("organization as facilitator")
- the organization being an additional, actual learning unit ("organization as learning unit")
- the organization being the end process (in contrast to the end product) that is dependent on learning and rests upon continuous learning to exist ("organization as end process").

Organization as facilitator may be identified through the arrangement of in-house courses, support of individual employees to take continuing education, the spreading of information and guidelines, and the arrangement of exchange of experiences. It may also be described through the climate for learning – how strongly and explicitly the organization opens for experimenting and spending time on learning activities.

Organization as learning unit may be identified through the distinction between the individual's learning as an agent and the organization's ability to gather, store, spread, and re-use that knowledge outside the individual. Argyris and Schön's (1978) work is generally considered important to understand this distinction. According to Robinson (2001) they belong to the normative strand on learning organizations using empirical studies to form theories on how organizations may improve by interventions. Robinson contrasts their approach to the descriptive strand. Argyris and Schön use manipulation of symbolic representations of organizational life, which is different from the non-symbolic, feedback-driven account of learning offered by other researchers. This points to, on one hand, the organization's effort to formalize and systemize knowledge and learning in systems, (symbolic) and on the other hand the organization's ability to facilitate direct transportation of knowledge from one individual to the next through direct dialogue (non-symbolic).

Organizations as end process is a more complex concept and understanding of organizational learning. It relates to teams and team development. The team is supposed to be able to solve problems and deal with emerging issues where team members can cover for each other. It requires strong focus on developing new skills and communication abilities. This

dynamic situation is a version of learning organizations that has a lot in common with organic structure – informal, decentralized, and non-hierarchical (Örtenblad, 2018, p. 153).

New perspectives on learning organizations consider learning as associated with the collective, more than the individual. New perspectives consider learning a social and cultural process, rather than a cognitive process. In addition, newer perspectives look at learning as context dependent, in contrast to the context independent learning dominating the traditional perspectives (Örtenblad, 2018, p. 153).

The focus on learning organizations is strongly associated with Peter Senge's seminal book *The fifth discipline: the art and practice of learning organizations* (Senge, 1990). Here he defines five disciplines ("component technologies") of learning organizations:

1. building a shared vision (common ideas that support selecting purposeful actions)
2. systems thinking (understanding how parts of the system work together)
3. mental models (how the individuals think about the world and how to take actions)
4. team learning (sharing experiences and learning together)
5. personal mastery (personal learning capacity and ability to act as a catalyst).

The shared vision is the organization's key tool to make sure individuals and units of the organization work to take the organization in the desired direction toward common goals and contribute to developing a culture of learning and sharing.

Systems thinking represents the need for a holistic view of the entire system. This underpins managers' and employees' ability to see that individual actions are correlated and dependent on others. This enables the understanding of patterns and improvement of systems.

Mental models concern individuals' understanding of who they are in the organizational context, what the business is about, and where they are collectively going in the future. The strongest organization is the one that is flexible and can adapt to changes by quickly developing new shared models.

Team learning is the key to accomplish excellent performance and a functioning organization. This is the discipline where personal mastery and shared vision come together. A safe and honest work environment where team members are colleagues and not rivals is essential for learning.

Personal mastery is not only about the individual competence but also the individual's clear vision of goals and perception of reality. The gap between vision and reality drives behavior in the direction of the vision if the truth about reality is shared. Empowering individuals to reach personal mastery is another key tool for the learning organization.

Senge's (1990) message is that these five disciplines cannot stand independently – they are interrelated. The learning organization needs to understand and apply them all to achieve a continuous learning process.

Nonaka (1991, p. 97) characterized knowledge-creating companies as places where "inventing new knowledge is not a specialized activity … it is a way of behaving, indeed, a way of being, in which everyone is a knowledge worker." Nonaka suggested that companies use metaphors and organizational redundancy to focus thinking, encourage dialogue, and make tacit, instinctively understood ideas explicit. In the governance context this highlights the importance of the development of an organizational culture that supports sharing of experiences and makes knowledge available for learning.

Returning to Argyris and Schön, in their book *Organizational learning II: theory, method and practice* (1996), they identify the two types of learning that might happen when there is a discrepancy between the intended result and the actual outcome of a task:

- single loop learning where you find a solution within the given frame of values, preconditions, standards, etc., to achieve the intended outcome
- double loop learning where assumptions and beliefs are questioned, and new thinking and methods can be found.

Single loop learning can be associated with "doing it right," and double loop learning may be associated with "doing the right thing" (Cartwright, 2002). Where doing it right concerns the operational short-term requirement to meet set standards, doing the right thing concerns the long-term improvement and organizational success. Here we may associate strategic perspectives. Both may be associated with governance, the setting of standards, and the defining of goals and strategies.

Garvin (1993) suggested that "learning organizations are skilled at five main activities: systematic problem-solving, experimentation with new approaches, learning from their own experience and history, learning from the experiences and best practices of others, and transferring knowledge quickly and efficiently throughout the organization. Each is accompanied by a distinctive mind-set, tool kit, and pattern of behavior."

Systematic problem-solving relies on scientific methods and facts, not guesswork. This requires access to data, systematic analysis, and effective ways of communicating knowledge. Experimentation involves systematic searching for and testing of new ideas and knowledge. It may be in the form of ongoing programs or one-of-a-kind demonstration projects. It is driven by opportunity, rather than problems, and requires incentives for risk-taking. Learning from past experiences requires systematic assessment of successes and failures. Records of lessons learned must be open and accessible.

Several definitions of lessons learned have been proposed – but a main point many authors have made is that the lesson (registered experience) is not learned until there is a change as a consequence (Rhodes & Dawson, 2013, p. 155). The lessons-learned literature thus focuses on systems for collecting (documenting) and connecting (sharing) experiences. The success factors of utilizing lessons learned are leadership, culture, roles, and responsibilities – and thus key issues in governance.

Lessons may come from successes and failures, but the failures more explicitly question why this outcome came to be. Knowledge from failures is often instrumental in achieving subsequent successes. A productive failure is one that leads to insight, whereas an unproductive success does not. Learning from others includes getting new perspectives from outside one's own immediate environment. Benchmarking is an example of systematic learning from others with a positive view. These studies should focus on practice – the way things are done – not the results. Transferring knowledge is important to make learning effective. It is far more helpful when shared among many than held in a few hands. There are many mechanisms by which knowledge can be transported including oral, written, and visual reports, visits, rotation programs, education, etc. They all have strengths and weaknesses (Garvin, 1993). In summary, the knowledge we have about learning organizations is that they need to master the craft of collecting, analyzing, and responding to facts about performance and how problems are solved. They need to make decisions based on facts, not guesswork. The learning organization

also needs to be able to learn from its failures and successes in a manner that helps the collective improve to a better version of itself, on the way toward a future vision that is shared in the organization. This involves actions on an organizational, team, and individual level.

Projects as Learning Organizations

Lessons learned is a recurring theme in project management (Rhodes & Dawson, 2013). This is due to the problems organizations face when trying to learn from experience made in fragmented project contexts and across organizations (Keegan & Turner, 2001). Divine and Zachry (2018) point out that project work is episodic, project professionals rarely have opportunity to reflect across projects to identify potential organizational, social, or tool-based contradictions affecting their work. Projects and project-based organizations do therefore have specific challenges.

Kerzner (2000) places continuous learning and improvement as the highest level of project management maturity. He says that "without 'discounted' lessons learned, a company can quickly revert from maturity to immaturity in project management. Knowledge is lost and past mistakes are repeated." Unlike the permanent organization, projects are action- and task-oriented and do not have the organizational mechanisms built in for learning. The outcome from learning processes is difficult to feed back to the permanent organization. Brady et al. (2002) hold that the uniqueness of projects is a challenge; their non-routine features hinder learning. Cooper et al. (2002) point out that the misguided belief that projects are too different to compare is a reason for a lack of learning in itself.

Williams (2008) demonstrated that there is a vital need to learn from one project to the next. He also shows that there are fundamental reasons why projects cannot easily create such learning. The temporary nature of projects is one aspect. The project exists only in a limited time and then the organization is dissolved. The consequence tends to be that the individuals take the new knowledge with them. The complexity of projects is another reason for lack of systematic learning. The complexity makes it hard to extract knowledge, typically from fragmented value chains and supply chains that span many different organizations.

Going back a few years earlier, Williams et al. (2001) observed that "even when post-project reviews are performed, there are no standard, structured, routine ways of analyzing projects to ensure that the organization can draw lessons and learn for future projects." That was then. Later, several methods were suggested; for example, Williams (2004) stated, "what are needed are simple, practical analysis methods that can be used routinely in post-project reviews to explicate how the project out-turn resulted and to identify the lessons which need to be learned," and demonstrated the practice in a case study.

There are also critical perspectives, like McAvoy and Butler (2009) calling projects reviews and post-mortems exercises in sense-making (p. 55), and mere feel-good exercises where the actors do not want to draw attention to failures (p. 57). They call for mindful use of post-mortem outcomes in project reviews and decision-making. Further, Pan et al. (2007) hold that the attribution theory literature serves as one of the few promising theoretical bases to explain why project post-mortems fail. They report a case study of a project post-mortem undertaken for an abandoned electronic procurement system project. They identify five antecedent conditions of attribution error: the presence of self-appointed mindsets, the general persistence of negative beliefs, memory decay, selective recall of project events, and the influence of power dynamics within the organization. All of these can reasonably explain the failure to learn the

right lessons from projects. Using attribution theory as a lens, they draw some conclusions as to how these challenges can be overcome.

Duffield and Whitty (2015) suggested a more specific model for organizational learning through projects. It is an adaptation of the so-called Swiss cheese model, which explains how project know-how can be distributed across an organizational network of elements such as individual learning, culture, social, technology, process, and infrastructure. The purpose is to conceptualize how to learn from past project experiences and successfully distribute project know-how across the organization.

In the continuing chapter we will use the term *ex ante* evaluations as similar to project reviews explained in the project management literature, and *ex post* evaluations as similar to post-mortems.

EVALUATION AS A SOURCE OF LEARNING

Evaluation is the systematic investigation of the feasibility of projects or other interventions. Unlike regular monitoring and control activities that follow the development continuously, it is an in-depth assessment which is done at defined points in the project's life cycle (*ex ante*, during implementation, or *ex post*). Evaluation requires certain expertise and application of scientific methods, while at the same time focusing on solving practical problems and being useful to project participants, decision-makers, and other stakeholders (Rossi et al., 2004).

Ex ante evaluation, in the form of project appraisal, is common in public as well as private projects, and is used to ensure that the right projects are selected and implemented efficiently. *Ex post* evaluation, on the other hand, is much rarer, especially in a strategic perspective. Many project owners do not know whether the intended effects were achieved, and whether assessments and forecasts were realistic or not. Worsley (2014) referred to *ex post* evaluations as "the weak link" in the assessment process for transport projects in Organisation for Economic Co-operation and Development (OECD) countries.

The purpose of evaluation is threefold, accountability, management, and learning. In this chapter, we discuss the learning aspect. Evaluations with a learning purpose should not just provide a "yes" or "no" answer to whether a project succeeded. Instead, they require an open-ended mandate to focus on and get a deeper understanding of causes and effects. Further, in a learning perspective, it is often more useful to study groups of several projects than only one project (Samset, 2003).

Evaluation can be an important source of learning. Samset (2015) refers to a study from the field of development aid, the sector that first started with project evaluations several decades ago. The study demonstrated that in a list comprising 19 learning sources, evaluation reports came in fifteenth. This indicates that learning potential is not realized. People learn first and foremost through their own experiences (they need to "make the mistakes themselves"). Secondly, they learn through direct communication, for example with colleagues talking about their own experiences. Reading evaluation reports may also provide useful input, but these tend to be forgotten more easily.

Therefore, it is important to plan for how to integrate evaluations in the organization's learning cycles, together with other sources of learning. Even if the evaluation is carried out by an independent party, there should be a plan for the involvement of affected parties, and a plan for how the results should be disseminated.

THE CASE: SYSTEMATIC *EX POST* EVALUATION OF PUBLIC PROJECTS IN NORWAY

On the backdrop of major cost overruns, delays, and benefit shortfalls in public projects in the 1990s, the Norwegian Ministry of Finance introduced a scheme in 2000 for *ex ante* evaluation of the country's largest public investment projects, the so-called quality assurance (QA) scheme (Volden & Samset, 2017). The intention was to ensure that projects were well planned with more realistic estimates of cost and benefits, before being presented to decision-makers for approval and funding. Twenty years later, about 300 projects have been exposed to this QA scheme.

The scheme has provided researchers with longitudinal data on the country's largest public projects, many of which are now completed and have now entered their operational phase. Since 2012, systematic *ex post* evaluations have been made of projects in their operational phase, under the auspices of the Concept Research Programme at the Norwegian University of Science and Technology. The evaluation framework and some results are presented below.

The Evaluation Scheme

A key question in an evaluation of a public investment project is whether the project was successful. Evaluation is thus a matter of determining the degree of success. Samset (2003) suggested a three-level framework, including the operational, tactical, and strategic level of project success. The three levels respectively concern (1) project delivery, (2) achievement of agreed goals, normally related to user benefits, and (3) societal effects.

We applied a broad, goal-oriented evaluation framework which covers all these three levels of success. It is based on a standardized set of five evaluation criteria, used by the UN and other institutions and development aid organizations, which has been endorsed by the OECD/Development Assistance Committee (OECD, 1991, 2002). We added cost–benefit efficiency as a sixth criterion – since cost–benefit analysis is mandatory in the front end of large Norwegian projects and has an important role in the QA scheme.

The six criteria are:

- *Efficiency* (operational level): This concerns project implementation and outputs in terms of cost, time, and quality, and how economically the project organization has converted inputs into outputs.
- *Effectiveness* (tactical level): This concerns whether the agreed outcome has been obtained and to what extent the project has contributed to this outcome.
- *Other impacts* (strategic level): This includes all consequences *beyond* the agreed outcome (side-effects), positive and negative, in the short and long term, for different stakeholders.
- *Relevance* (strategic level): A project is relevant if there is a *need* for what the project delivers. Need is seen in relation to national political priorities, measures of demand, and stakeholders' preferences. Conflicts of interest should be brought to light.
- *Sustainability* (strategic level): A project is sustainable if its benefits are likely to persist throughout the lifetime of the resulting infrastructure. This usually requires that the total impacts (financial, environmental, and social) are acceptable in the long run.
- *Cost–benefit efficiency* (strategic level): This should be measured in terms of total willingness to pay in relation to cost, or secondarily in terms of outcome in relation to cost.

More details on the evaluation framework and methodology are presented in Volden (2018).

The evaluation model is broad to provide an *overall* picture of the degree of public project success. The purpose is to learn and improve in terms of "doing the project right" as well as "doing the right project": single loop learning as well as double loop learning.

At the same time, it is a fairly simple scheme that only includes six generic evaluation criteria, which should be possible to meet for all types of projects. With budget limitations, we cannot be too ambitious regarding the methodological rigor when responding to each criterion. We see it as more important that evaluations are carried out for a significant number of projects.

A main feature is that all evaluations should apply the same format. This is to facilitate learning, since it allows us to draw lessons from aggregates of evaluations. Not only did we apply the same six criteria in all project evaluations, but we also aggregated the results by setting a score between 1 and 6 for each criterion in all projects using a predefined scale. This was to facilitate comparison of achieved success across projects.

It should be noted that all the evaluations were done several years after the projects were completed (i.e., post-mortem). Thus, these are *summative* evaluations, not *formative* ones. There was no active project team in place at the time of evaluation, and the purpose was not to improve performance of an ongoing project. Instead, the purpose was for all those involved to learn in order to *improve future project practices*.

Evaluation Results

Between 2012 and 2021, 29 projects have been evaluated according to this scheme: 10 road projects, seven buildings, four railway projects, six information and communication technology (ICT) projects, and two defense projects.

They constitute about half of all the projects that have undergone QA and have been *in operation for at least five years*. The Concept Research Programme has been responsible for the evaluations, but a different evaluation team was established in each case. The evaluation team could not have any conflict of interest in relation to the project. In all, 19 research and development organizations and consultancies have been involved.

The evaluation team reviewed and, if necessary, adjusted the project's internal logic or goal hierarchy. The team then operationalized each of the six criteria, and then carried out the evaluation by collecting and analyzing data, combining different data sources and methods. They interviewed participants from the original project team, the public agency being responsible for the projects, its parent ministry, and other stakeholders. Finally, the team summarized its assessment for each criterion by setting a score between 1 and 6. The final report was made public and distributed.

The findings in terms of evaluation scores are presented in Table 20.1. Score 1–2 are shown as numbers in italics, score 3–4 as plain text, and score 5–6 in bold fonts. Reservations should be made concerning the accuracy of the scores. Unavoidably, a certain degree of subjective assessment must be accepted from the evaluator's side.

The overall picture of performance is quite positive, with average scores between 4 and 5 for all criteria. Only eight out of 29 projects had cost overruns, and the scores concerning efficiency were generally high. The weakest result concerns cost–benefit efficiency. Evaluators noted that this was the most challenging criterion to assess. A general problem for evaluators was the lack of data on key variables related to effects (for example, traffic volumes,

Table 20.1   Results from 29 ex post evaluations in the operational phase of public projects in Norway

Project	Sector	Evaluation criterion (level of success)					
		Efficiency (OPERATIONAL)	Effectiveness (TACTICAL)	Other impacts (STRATEGIC)	Relevance (STRATEGIC)	Sustainability (STRATEGIC)	Benefit–cost eff. (STRATEGIC)
1	Road	3	5	4	5	5	6
2	Railway	5	3	4	5	5	3
3	Defense	4	4	4	3	2	2
4	Building	6	4	4	5	5	4
5	Road	4	5	4	3	4	3
6	Road	5	6	5	5	5	6
7	Road	5	6	5	4	5	5
8	ICT	5	6	4	3	5	3
9	Building	6	4	5	5	5	3
10	Railway	4	4	4	5	5	3
11	ICT	5	5	5	5	6	5
12	Road	4	5	3	5	4	4
13	Defense	5	5	5	6	5	5
14	Building	5	4	3	4	5	4
15	Road	5	5	5	5	5	6
16	Building	5	5	6	5	4	4
17	Building	5	4	5	4	5	4
18	Road	4	5	4	5	4	6
19	Road	5	5	4	5	4	6
20	Railway	4	3	4	4	4	2

(*Continued*)

Table 20.1   (*Continued*)

Project	Sector	Evaluation criterion (level of success)					
		Efficiency (OPERATIONAL)	**Effectiveness** (TACTICAL)	**Other impacts** (STRATEGIC)	**Relevance** (STRATEGIC)	**Sustainability** (STRATEGIC)	**Benefit–cost eff.** (STRATEGIC)
21	Railway	4	3	3	3	4	2
22	Building	2	**5**	**5**	**5**	3	4
23	Road	4	4	3	3	3	3
24	Road	3	**5**	**5**	**5**	**5**	**6**
25	ICT	4	4	4	**5**	**5**	**6**
26	ICT	**5**	4	3	**5**	4	4
27	ICT	*1*	2	2	3	2	*1*
28	ICT	4	4	**5**	**5**	**5**	3
29	Building	**5**	4	**5**	4	**6**	**5**
Average	score	4.3	4.4	4.3	4.5	4.5	4.1

Note: Good results **5–6**; acceptable results 3–4; bad results 1–2

punctuality, and accident levels in the case of transport projects). Project appraisals and *ex ante* data of better quality would have made *ex post* evaluation easier.

Generally, there were few projects which could be considered outright successes or failures (except four projects that received a green light across the board). The majority showed more ambiguous results, being successful according to some perspectives and less successful in others.

There is some correlation between scores on the three levels of success, which indicates that a project that is well planned and governed in one perspective has a greater chance of succeeding in other perspectives as well.

The results indicate some interesting sectoral differences, although care should be taken when comparing scores across sectors since different project types may have very different goals and levels of ambition. *Building projects* scored very well on operational success, but slightly lower tactically and strategically. *Railway projects* were closely aligned with government strategies for a "green shift" in transport, and thus were considered relevant and sustainable. However, they scored very low on cost–benefit efficiency and on effectiveness. *Road projects* scored high on most criteria, but somewhat lower on efficiency, and on other impacts due to negative (environmental) side-effects that had not been properly dealt with.

The promising results indicate that many project managers, owners, and funders do a good job in ensuring public project success, but also that there is potential for improvement in almost all projects, and that there is much to learn across projects and sectors.

Some learning examples from an operational perspective include cost overruns, which could often be explained by a tight construction market – recommendations for agencies are to ensure good timing of tenders, adjust the contract scope to maximize competition, and communicate actively with the marked in advance; geology – the ground conditions need to be properly investigated before one starts to build; and the need to involve user groups and operating personnel in the early stages of projects. Many projects did this, but some that did not had to make costly changes and adjustments after the project was completed, in order to ensure good operation and functionality. A fourth factor to highlight is that humans are of crucial importance, and that disputes in projects can be devastating. This should be taken more seriously in projects. In many projects there is a need to replace personnel.

Evaluators also find that the work done on benefits management is often unsatisfactory, resulting in lower scores than necessary on effectiveness and cost–benefit efficiency. Some of the projects had very ambitious goals that required more than the physical infrastructure to be achieved. Still, few organizations had benefit-realization plans with assigned responsibility to benefit owners or project owners. Evaluators generally recommend that more projects take benefits seriously – "they do not occur by themselves."

Some projects had positive side-effects, resulting in high scores on other impacts and sustainability. Generally, the building projects did best in terms of maximizing positive side-effects, typically depending on the location of the building. Others had negative side-effects (on the environment or the local community) which could have been avoided, and still others missed the opportunity of positive side-effects (for example, road projects in rural areas which could have been exploited in a local/regional growth strategy). Both positive and negative side-effects could have been avoided with earlier stakeholder involvement, and in many cases with use of compensatory and supplementary measures.

Overall, most projects scored high on relevance, indicating that there was a need for the service that the project generated. However, the relevance criterion is multidimensional, and

some projects involved conflicts of interest. Further, in some cases, evaluators noted that other, more relevant solutions to the problem existed (at least more cost-efficient ones). This should be seen in relation to the often-low scores on cost–benefit efficiency. For example, some of the ICT projects studied involved the development of new, complex solutions, although simpler, off-the-shelf systems were available. Not all public projects can or should be "profitable," but one should at least consider whether a simpler solution, still with acceptable goal achievement, would substantially improve value for money.

Use of the Evaluations

The purpose of doing *ex post* evaluations is, ultimately, to improve the success of future projects. The organizations involved should compare all the evaluations of their own projects, compare with experiences from projects in other sectors, and look for useful lessons.

For learning to happen, several conditions must be met. The evaluations ought to be of good quality, and the results and recommendations must also be perceived, understood, and used. Several steps were taken by the researchers to ensure this. Admittedly, these evaluations are somewhat "rapid," and the scores sometimes uncertain, as is not uncommon in evaluations. Acceptable quality was ensured in various ways, most importantly by using *triangulation* (i.e., a wide range of sources and methods) as a key strategy for data collection and analysis. Further, we made guidelines for score-setting available to the evaluators, to promote calibration of assessments. We also made efforts to involve the affected parties (ministries and agencies) during evaluations, and we presented and discussed the findings with them afterwards. All the reports had executive summaries in which results and key lessons were summarized.

Evaluations are not always used as intended (Dahler-Larsen, 2012; Samset, 2015). *Wrong use* of evaluations can be a problem. For example, instrumental use of evaluation results to reward or punish those involved may create barriers to truth-speaking and real learning. Even more common than wrong use is *non-use* of evaluations, which has been a topic within the evaluation literature since the 1970s. It is well documented that many evaluations are more ritual and symbolic than truth-seeking. Possible explanations can be related to the quality of evaluations, low perceived relevance to the target group, lack of standardization (making it difficult to compare results across projects), and explanations related to power and politics (unwillingness to use the results).

As noted by Scriven (2015), the use or non-use of an evaluation's results cannot be regarded as a quality criterion of the evaluation itself. Even evaluations of high quality, with clear communication, right timing, and user involvement, may be disregarded due to bad management (Scriven, 2015). This is the case not least in the public sector, where horse trading and other political considerations may be emphasized more than facts and knowledge in decision processes. On the other hand, there is no doubt that a good evaluation will have a higher probability of being used than a poor evaluation, all else being equal.

A separate survey was conducted by the Concept Research Programme to investigate the perceived benefits and use of evaluations (Bukkestein et al., 2020). Interviews were conducted with senior project owners and employees involved in project planning and appraisal in affected ministries and agencies.

A key finding was that the target group had varying knowledge of the evaluations concerning their own sector. Those who had been involved and were aware of the evaluations often showed considerable interest in them. Their colleagues were sometimes not even aware of the

evaluations. Thus, evaluations did contribute to individual learning, but to a lesser extent to organizational learning. This is a well-known challenge in project-based organizations, and applies to all sources of learning, not only evaluations. It is often explained by the temporality as well as the uniqueness of projects.

A complicating factor in this case was that the evaluations were initiated and carried out by an external party (a research program). The findings of external evaluations normally have higher credibility than those of internal ones, especially for stakeholders and other parties outside the organization. On the other hand, external evaluations may generate less interest in the organization itself. Interestingly, the same survey found that very few (hardly any) evaluations were initiated by the ministries and agencies. Project appraisal and planning was thorough and systematic, and involved external QA according to the Ministry of Finance's scheme. Only a few evaluations had been carried out *ex post*. These evaluations were directed on operational success and had essentially a control function, rather than learning as their focus.

Bukkestein et al. (2020) provided a set of recommendations on how to improve the use and benefits of evaluations, the most important being:

- involve the relevant ministry and agency earlier and more systematically in the evaluation process
- sort the findings and recommendations more clearly by target group; normally, operational project success concerns the agency most, tactical success concerns both agency and ministry, and strategic success concerns the ministry most
- focus less on the scores themselves and more on the learning points, when the report is written
- focus more on dissemination of the results; this should be a task for both the Concept Research Programme and the organizations involved
- organizations should take responsibility for some evaluation activities themselves, to supplement the work done by external parties.

CONCLUDING REMARKS

The implemented governance framework evaluated in this case shows signs of focusing quality at entry over quality at exit. In addition, the focus on learning and long-term improvement of the governance could benefit from more attention.

Project evaluations for learning purposes should be conducted routinely and according to a common evaluation framework, rather than as an ad hoc activity. In that way, it will be easier to draw lessons from successes as well as failures. Evaluating more projects with a simple and standardized evaluation method is more influential than deep analysis of a few single projects. Looking for patterns and improvement potential at system level can make a real difference if they result in learning across projects and organizations. Unfortunately, evaluations are often initiated only after something has gone wrong, and tend to focus more on the control aspect than learning.

Going back to Örtenblad's (2018) distinction between "organized learning" and "organizational learning" mentioned in the introduction, we believe that *ex post* evaluation can contribute to both. First, by making evaluation part of the organization's formalized governance framework, one may ensure that evaluations are actually conducted, and the results collected

and disseminated throughout the organization (i.e., organized learning). Secondly, the organization should strive to build a climate for learning, from one's own projects as well as others', including willingness to use the evaluation results to question old practice in light of new knowledge (i.e., organizational learning).

The active use of *ex post* evaluations opens up double loop learning (Argyris & Schön, 1996), which supplements the single loop learning that might follow from control and measurement activities in project management. This also equips the organization for mature project management according to Kerzner (2000). As the evaluation models reported here show, the challenges reported by Brady et al. (2002) can be overcome. The uniqueness of projects is not an excuse to refrain from trying. The experience reported here also rejects theories that claim *ex post* evaluations should fail, like "feel good exercises unwilling to draw attention to failure" (McAvoy & Butler, 2009). However, the challenges represented by fragmentation and the effect of time on memory and attribution remain.

One potential remedy to help against these challenges is making sure facts are measured and documented in a format that makes them available for later evaluation and support for planning and decision-making. Having entered the digital era, this should not be an unrealistic expectation. Experience so far does not give promises of a quick fix. Digitization has promised a lot more than it has delivered so far. Still, we have not given up on such future perspectives. Digital systems do hold the potential for supporting broad learning and general improvement.

The building of a database with actual results from completed projects may contribute to more realistic cost figures, as well as more realistic estimates of benefits. This will support more precise predictions, realistic decisions, and robust project planning. It may even facilitate futuristic systems that may include artificial intelligence and machine learning, estimation methods like reference class forecasting and case based reasoning, i.e., ways of predicting the future by looking at similar past situations and their outcomes. The value of such a database will increase over time with an increasing number of projects.

Society should expect that decision-makers become equipped with better and more empirically based analyses supporting better use of public finances and sustainable development. Organizations at all levels of the project environment should welcome the effect of learning across projects and over time that would improve performance toward increased competitive strength and improved reputation. One motor that makes this development possible is the systematic use of effective evaluations in the front end, but also *ex post* after the project's resulting infrastructure has been in operation for a few years. Embedding this good practice in the governance frameworks helps spread the effect across the public sector.

REFERENCES

Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective.* Addison-Wesley.

Argyris, C., & Schön, D. A. (1996). *Organizational learning II: Theory, method and practice.* Addison-Wesley.

Brady, T., Marshall, N., Prencipe, A., & Tell, F. (2002). Making sense of learning landscapes in project-based organisations. 3rd European Conference on Organizational Learning, Knowledge and Capabilities hosted by ALBA, Athens, Greece, 5–6 April.

Bukkestein, I., Welde, M., & Volden, G. H. (2020). *Bruk og nytte av etterevalueringer av prosjekter.* Concept working paper (pp. 2020–2023). https://www.researchgate.net/profile/Morten-Welde/

publication/348416220_Bruk_og_nytte_av_etterevalueringer_av_prosjekter/links/5ffdb836299bf14 0888cf1d6/Bruk-og-nytte-av-etterevalueringer-av-prosjekter.pdf

Cartwright, S. (2002). Double-loop learning: A concept and process for leadership educators. *Journal of Leadership Education*, *1*(1), 68–71.

Cooper, K. G., Lyneis, J. M., & Bryant, B. J. (2002). Learning to learn from past to future. *International Journal of Project Management*, *20*(3), 213–219.

Dahler-Larsen, P. (2012). *The evaluation society*. Stanford Business Books.

Divine, D., & Zachry, M. (2018). Project management, contradictions, and textualized activity: Supporting reflection in project based organizations. *Technical Communication*, *65*(2), 194–209.

Duffield, S., & Whitty, S. J. (2015). Developing a systemic lessons learned knowledge model for organisational learning through projects. *International Journal of Project Management*, *33*(2), 311–324.

Garvin, D. (1993, July–August). Building a learning organization. *Harvard Business Review*, 78–91.

Greenwood, J. (2003). The role of reflection in single and double loop learning. *Journal of Advanced Nursing*, *27*(5), 1048–1053.

Keegan, A., & Turner, J. (2001). Quantity versus quality in project based learning and continuous improvement. Management Learning, *32*(1), 77–98.

Kerzner, H. (2000). *Applied project management: Best practices on implementation*. New York: Wiley.

McAvoy, J., & Butler, T. (2009). Project post-mortems: Mindless mismanagement of agreement. *Journal of Decision Systems*, *18*(1), 53–73. https://doi.org/10.3166/jds.18.53-73

Nonaka, I. (1991, November–December). The knowledge-creating company. *Harvard Business Review*, *85*(7/8), 162.

OECD. (1991). *DAC principles for evaluation of development assistance*. OECD, Development Assistance Committee.

OECD. (2002). *Glossary of key terms in evaluation and results based management*. OECD, Development Assistance Committee.

Örtenblad, A. (2018). What does "learning organization" mean? Editorial. *Learning Organization*, *25*(3), 150–158.

Pan, G., Pan, S. L., & Newman, M. (2007, December). Information systems project post-mortems: Insights from an attribution perspective. *Journal of the American Society for Information Science and Technology*, *58*(14), 2255–2268.

Rhodes, L., & Dawson, R. (2013). Lessons learned from lessons learned. *Knowledge and Process Management*, *20*(3), 154–160.

Robinson, V. M. J. (2001). Descriptive and normative research on organizational learning: Locating the contribution of Argyris and Schön. *International Journal of Educational Management*, *15*(2), 58–67. https://doi.org/10.1108/EUM0000000005395

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). SAGE.

Samset, K. (2015). Reell eller rituell evaluering. *Stat og Styring*, *3*, 6–9.

Samset, K. F. (2003). *Project evaluation: Making projects succeed*. Tapir Academic Press.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Scriven, M. (2015). *Key evaluation checklist (KEC)*. http://michaelscriven.info/papersandpublications .html.

Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. Doubleday.

Volden, G. H. (2018). Public project success as seen in a broad perspective: Lessons from a meta-evaluation of 20 infrastructure projects in Norway. *Evaluation and Program Planning*, *69*, 109–117.

Volden, G. H., & Samset, K. (2017). Quality assurance in megaproject management: The Norwegian way. In B. Flyvbjerg (Ed.), *The Oxford handbook of megaproject management* (pp. 406–427). Oxford University Press.

Williams, T. M. (2004). Identifying the hard lessons from projects – Easily. *International Journal of Project Management*, *22*(4), 273–279.

Williams, T. M. (2008, May). How do organizations learn lessons from projects – And do they? *IEEE Transactions on Engineering Management*, *55*(2), 248–266.

Williams, T. M., Ackermann, F., Eden, C., & Howick, S. (2001, November). The use of project post-mortems. Presented at the Project Management Institute Annual Symposium.

Worsley, T. (2014). *Ex post assessment of transport investments and policy interventions: Prerequisites for ex-post assessments and methodological challenges. ITS roundtable summary and conclusions.* International transport foundation discussion paper 2014–19. https://www.econstor.eu/bitstream/10419/109153/1/818314141.pdf