IOP Conf. Series: Earth and Environmental Science **1389** (2024) 012025 doi:10.1088/1755-1315/1389/1/012025

Lessons from ex-post evaluation of government investment projects

Morten Welde[1*] **and Gro Holst Volden** [1]

[1,]Department of Civil and Environmental Engineering, NTNU, P.O. box 8900, NO-7491 Trondheim, Norway

[*]E-mail corresponding author: morten.welde@ntnu.no

Abstract. Although public goods and services are increasingly delivered through projects, many studies have demonstrated that the success of projects in different sectors varies. In Norway, large government projects are subjected to extensive ex-ante appraisal with external quality assurance before budgets can be allocated, and each year, a sample of projects is also evaluated ex-post under the auspices of a research programme. The ex-ante processes and the evaluation of projects ex-post are unusual by international standards, which thus provides an interesting foundation for exploring how the knowledge gained from ex-post evaluations is used to improve the planning and delivery of future projects. This paper presents a study of how Norwegian ministries and agencies use ex-post evaluations of projects within their respective portfolios of responsibilities. Based on interviews with representatives of the target group for the evaluations, the authors examine whether the evaluations improve the appraisal and delivery of subsequent projects. The findings reveal that the use of the results has been limited and that many target group members are unaware of the evaluation. The authors suggest some changes to the evaluation process that might increase the use of ex-post evaluations.

1. Introduction

Projects have become the dominant means of organising work and are increasingly used to deliver public goods, services and policy [1]. However, large government-funded projects have a grim track record of delivery. Cost overruns, delays and benefit shortfalls are common [2]. One of the fundamental mechanisms behind the success or failure of such projects is the ability to learn from completed projects and to use that knowledge to improve the performance of future projects. Ex-post evaluation is a tool to assess whether a specific intervention was justified and whether it worked as expected in achieving its objectives and why. It can be used to check whether projects delivered the expected benefits and to learn which projects do better and which do worse than expected and why [3]. Ex-post evaluation links the projects' forecasts, estimates and expectations with their actual results and impacts. The ability to learn from past performance through evaluation is thus crucial.

In recent years, the project management literature has paid increasing attention to the front-end phase of projects and its importance for project success ([4][5][6][7][8]). The project's front-end phase is the stage from the first initiative until the project receives a formal budget, and construction or execution can start. In projects that experience problems with cost overruns, delays and poor value for money, the source of the problems is often found in the front-end phase[7]. Therefore, we must have tools and systems to distinguish broadly between good and bad projects early in the appraisals. To avoid optimism bias, these systems should include independent reviews of decision documents [4].

Ex-ante appraisal of project success is well developed, and most countries use a lot of resources to estimate costs and forecast, the likely user effects and societal impacts of projects. However, when a project is completed and the results are put into operation, many questions may remain: Did the project deliver its results as planned? Were the benefits as expected? Did the project contribute to wider impacts? Knowledge of the actual results is important for control purposes but even more important for learning purposes, as well as to increase the quality of future ex-ante appraisal and project delivery.

Nevertheless, ex-post evaluations of projects are rarely conducted ([9][10]). Therefore, knowledge of projects' actual value for money, effectiveness, and sustainability may be limited, and why the results turned out the way they did remains unknown. This contrasts with the advice of many scholars to conduct ex-post evaluations to improve appraisal tools and models, reduce inaccuracies in estimations of costs and benefits, and improve project implementation practices ([11][12][13]). However, conducting ex-post evaluations is not enough for practices to improve – the lessons from evaluations must also be known, recognised as relevant, and used.

This paper explores the use and benefits of systematic ex-post evaluations conducted by a research programme funded by the Norwegian Ministry of Finance. Norway provides an interesting case for examining the use of evaluations for two reasons. First, its extensive processes for ex-ante appraisal, including external quality assurance in two stages, provide detailed information against which the ex-post results can be mapped. Large government projects are planned and delivered within a common governance framework, allowing comparisons between projects and sectors. Second, the fact that ex-post evaluations are carried out systematically for learning and improvement purposes by a party independent of the projects is uncommon by international standards. Although the number of projects evaluated yearly is low, the evaluations are carried out using the same evaluation model. Over time, the increased sample of evaluated projects has allowed for increased knowledge of project performance. Within this context, it is relevant to study whether this knowledge is actively used by the ministries and agencies responsible for the projects.

In this paper, we thus investigate if the ministries and agencies responsible for the planning and delivery of large projects in Norway:

- Are aware of the evaluations and their results.
- Use the results in current and future projects.
- Carry out evaluations themselves.

We also explore what prerequisites must be met or barriers overcome to ensure that post-evaluations are useful.

Accordingly, this paper aims to contribute to the literature on evaluation and its use. The results should be relevant to organisations already performing ex-post evaluations and those planning to do so.

We begin by discussing some findings from the literature on evaluation use. We discuss different types of use and some of the main reasons for not using evaluations. After that, we present the data and the evaluation framework used. In the results section, we report the findings from interviews, a workshop, and document studies. Finally, we discuss the implications of the findings and propose a series of measures that could be taken to increase the likelihood of evaluation results being relevant and useful for those whose projects are evaluated.

2. Ex-ante and ex-post evaluation in Norway

In 2000, the Norwegian Ministry of Finance introduced a mandatory quality-at-entry scheme for the largest public projects to ensure they were well-prepared [14].

The quality assurance scheme was introduced following unfortunate experiences with some large public investment projects involving large cost overruns, delays, and low benefit-cost efficiency.

The scheme comprises two extensive appraisal studies, followed by external QA reviews in an investment project's planning process:

- QA1 – Quality assurance of the choice of concept before a Cabinet decision to start a pre-project.
- QA2 – Quality assurance of the management base and cost estimates before the project is submitted to Parliament for approval and funding.

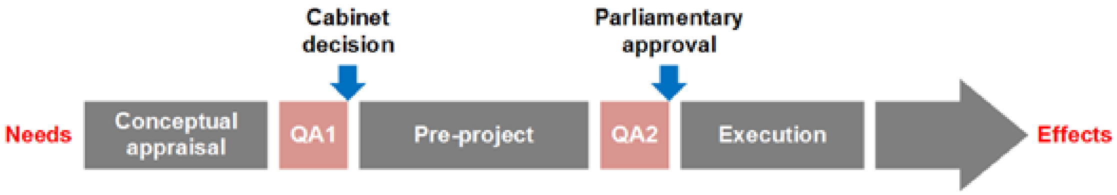

Figure 1: The Norwegian scheme for quality assurance of major public investment projects (the QA scheme)

The QAs are performed by a consortium of external consultants pre-approved by the Ministry of Finance. The QA scheme, as such, applies to the projects' planning phase and does not include an explicit requirement for ex-post evaluation of completed projects. The regulations for financial management in central government include only a very general requirement for ministries and agencies to "ensure that evaluations are carried out to obtain information on efficiency, goal achievement and results within their area of responsibility and activities" ([15] p. 16).

To gain knowledge about the results of projects in the QA scheme, the Ministry of Finance has funded a research programme since 2002 to follow the projects and perform ex-post evaluations: The Concept Research Programme. Concept applies a generic six-criteria evaluation framework inspired by a framework frequently used to evaluate development assistance projects. The criteria are (1) efficiency, (2) effectiveness, (3) other impacts, (4) relevance, (5) sustainability, and (6) benefit-cost efficiency. The model is inspired by a model originally developed by the OECD Development Assistance Committee [16]. The 'efficiency' criterion measures project success from an operational perspective (project management success), whereas the other five criteria measure tactical and strategic success (i.e., how the project contributes to user- and societal benefits). Since 2023, between 2-4 projects each year have been evaluated. Each time, an evaluation team is established with 3-4 persons (researchers or hired consultants) independent of the project to be evaluated.

Volden [17] presented the evaluation model and some preliminary findings; for a description of its use in the transport sector, see Welde and Volden [18]. The results indicate that most projects were successful, especially concerning cost performance. However, the evaluations also demonstrate considerable room for improvement, for example, in terms of benefits realisation and value for money.

In the long run, the benefits of the evaluations and similar ones will depend on whether the results are known, understood, and used in planning new projects. Unfortunately, evaluations are not always used for their intended purpose, and their function might be more ritual than real ([19][20]).

In the evaluations discussed here, the organisations whose projects are evaluated do not participate in the evaluation themselves, apart from as informants, and the knowledge acquired by the evaluators may disappear from the organisation after the evaluations have been published.

This is a general challenge with externally conducted evaluations. Therefore, it is important to know what is needed for an evaluation to be used in the evaluated organisation and for the evaluation results to be considered when planning new projects.

3. Literature review

The use or non-use of evaluation has been a topic within the evaluation literature since the 1970s when concerns about lack of use led to numerous empirical studies being conducted ([21][22][23][24]). Several scholars claim that this is one of the most extensively studied areas within evaluation as a subject ([25][26]). Despite this, the problem of evaluation use and non-use remains unsolved [27]. While some argue that the main problem is that evaluations are not carried out at all [10], in this paper, we focus on the lack of use of the results of completed evaluations.

3.1. Different types of evaluation use

An evaluation and its results may come to use in several different ways. Vedung [28] proposed six types of evaluation use:

- *Instrumental:* An evaluation that provides direct input, which results in immediate action.
- *Conceptual*: An evaluation that influences how potential users think (i.e. leads to their "enlightenment"). This type of use may stimulate thinking but will not directly and immediately transform into decisions and actions.
- *Legitimising*: An evaluation is employed to endorse previously established positions without changing them. "Persuasive", "argumentative", and "symbolic" are other terms for this use.
- *Tactical:* Initiating an evaluation as an excuse for inaction.
- *Ritual:* An evaluation that is justified as one of a series of acts that is always done in a particular situation and performed in the same way each time.

Vedung [28] argued that all types of uses may refer both to evaluation products and to evaluation processes, whereby product use is the employment of evaluation findings from final reporting, and process use is the utilisation of any evaluation activity before final reporting.

Different types of evaluations may have different objectives. Steering and control purposes are often most prominent in pre- and mid-term evaluations. In such cases, one would like to see a direct instrumental use so that the evaluation results lead to immediate action (e.g., terminating the project). The lack of use of results from ex-ante evaluations in decision-making has been documented in several studies. For example, Eliasson et al. [29] showed that ex-ante cost-benefit analysis of Norwegian road projects did not affect which projects were selected for implementation in the Norwegian transport plan for 2014–2023. In a study of the Danish municipal sector, Dahler-Larsen [30] found that less than 1% of 600 evaluations had led to the termination of the evaluated activity. Although termination rates were low, the findings suggested that the evaluations did lead to incremental changes, as almost 80% of the evaluated activities were adjusted.

The main purpose of ex-post evaluations is to learn how past failures and successes can contribute to making better choices and improved implementation in future projects (i.e., the main objective is conceptual use). There is great potential to learn from both those who succeed and those who fail so that estimates, forecasts, and tools for use in future project appraisal can be improved.

In practice, evaluation is not necessarily considered an important source of learning. In a review of Norwegian state enterprises' use of evaluation, one of the main findings was the lack of follow-up of evaluation results [31]. At worst, evaluation becomes a ritual or symbolic act rather than a truth-seeking process. Samset [20] illustrated this aptly by describing the results from a 1993 meta-evaluation of the use of evaluation. Caseworkers and decision-makers from a development aid administration were interviewed about which arenas they considered most important for learning. Of 19 arenas, evaluation reports were in 15th place and thus considered among the least important sources of learning, despite the extensive scope of evaluations in the field of development aid.

3.2. What causes the lack of use?

The literature points to several reasons why the results of previous evaluations are not being used. The lack of use is often the result of many factors. Some of the factors can be categorised as follows.

Poor quality of the evaluation report. The lack of use of evaluations can sometimes be traced back to poor data quality, methods, analyses, or conclusions. Poor quality of available data and lack of access to data are recurring problems, and the challenge is often greater for ex-post evaluations than for ex-ante evaluations, especially in the absence of a baseline assessment ([32][33]). Ledermann [34] found that evaluation quality, as perceived by the decision-makers, appeared to matter in most contexts. However, Christie [26] found that high standards of methodological rigour were among the less important factors in whether evaluations were used. This indicates that perceived quality is the key factor.

Lack of relevance for the target group. If the results of an evaluation and its recommendations are to be considered, they need to be perceived as relevant to those who are evaluated. This may be a particularly big problem when an evaluation is conducted by external actors [26]. The problem with relevance may be an argument for organisations to conduct evaluations internally. On the other hand, this may compromise the “critical eye,” which is crucial for the trustworthiness and quality of evaluations. The pros and cons of internal vs external evaluation should always be discussed, but pure internal evaluations are rarely recommended.

Lack of anchoring and participation by those who are evaluated. Most people learn more from first-hand experience than from reading reports. Oxera [35] examined ex-post evaluations of trunk road schemes and found no mechanisms for drawing together and addressing different interests. The range of interests was limited to a few officials, and many consultees were unaware of the evaluations. They, therefore, concluded that the dissemination of findings was ineffective. Johnson et al. [36] reviewed the empirical literature on the use of evaluations from 1986 to 2005. They found that interaction and communication between those evaluated and those who completed the evaluations were decisive for whether the results were utilised. Similar results were found by Berriet-Solliec et al. [37], Christie [26], and Milzow et al. [38].

Lack of standardisation. An evaluation can be good and relevant, but as much as there is to learn from one evaluation, there is more to learn from several evaluations of the same phenomenon. However, if lessons are to be drawn from several evaluations, their results must be comparable. Nicolaisen and Driscoll [12] examined schemes for ex-post evaluation in the transport sector in different countries. They concluded that the lack of standardised methods weakened the possibility for comparison and, thus, the learning potential.

Lack of integration into the organisation’s management and governance cycles. Evaluations are performed in addition to ongoing monitoring and control activities. This entails a risk that evaluation will become ad hoc and that there will not be a system to ensure that the results will be included as a basis for assessing or planning the next project. The Norwegian Agency for Public and Financial Management (DFØ) has created a guide to help companies link evaluations to their management processes [39]. Bastøe [40] argued that there is a need for evaluation activities to be institutionalised so that it will be clear what should be evaluated, how evaluations should be carried out, and how they should be used and followed up. Bastø stated that DFØ should be clearer in its recommendations on how evaluations should coexist with other sources of information and how an evaluation should be used and coordinated to ensure a comprehensive and balanced knowledge base.

Political reasons and unwillingness to learn. Regardless of how good an evaluation is and how well the arrangements for its use are, decision-makers will not be guaranteed to learn. As Samset [20] pointed out, evaluations intervene in a reality of rational and political choices that do not always coincide. For the same reason, Scriven [41] considered that a distinction should be made between the usability of evaluations and actual use and, thus, that the quality of an evaluation cannot be assessed based on whether it is used.

4. Data and empirical strategy

The purpose of the study presented in this paper was to provide knowledge about the use and benefits of ex-post evaluation of government investment projects conducted under the auspices of the Concept research programme, particularly the conditions that must be satisfied for the evaluations to be useful. Through interviews and document studies, we aimed to reveal whether evaluations were used and, if not, to provide advice on how they could be tailored to meet the needs of project-based public organisations better and how the results of evaluations could be disseminated more effectively.

4.1. The evaluated projects

The starting point for our study was 26 ex-post evaluations carried out by Concept [42]. We investigated whether the evaluation reports had been disseminated within the organisations responsible for the projects and whether the results had been used in the planning and appraisal of new projects. All 26 evaluations were of large government investment projects subjected to external QA2. Table 1 presents the main characteristics of the projects that were evaluated.

Table 1. The evaluated projects

Sector	No. of projects	Completed year(s)	Evaluated years	Average final cost in EUR million (nominal values)
Road	10	2005–2013	2012–2019	125
Buildings	6	2005–2011	2012–2018	130
Rail	5	2005–2011	2012–2019	195
ICT	3	2010–2013	2014–2019	70
Defence	2	2012	2012–2015	320

4.2. Data collection

The study was based on a combination of semi-structured interviews and document analysis. We interviewed 30 people, 19 from ministries and agencies subject to the QA scheme and 11 from other public organisations with knowledge and experience of evaluations. Most interviewees were senior project owners or employees involved in project planning and appraisal.

The interviews lasted 90 minutes on average. They were based on a semi-structured interview guide with three main topics to allow for comparisons of the interviewees' answers: (1) knowledge of evaluations, (2) use of the results in current and future projects, and (3) the extent to which organisations carry out evaluations themselves. The third question was asked to explore whether a reason for the non-use of the Concept evaluations could be that the organisations preferred internal evaluations.

At least two researchers were present during each interview. An additional source of material for the study was a workshop on using evaluation results. A total of 15 people from ministries, agencies, and consultancies involved in external QA participated in the workshop.

To supplement the interviews, we examined the extent to which project appraisals and their accompanying QA1 reports explicitly used or otherwise referred to results from Concept's evaluations of similar projects. Additionally, we examined publicly available information about the agencies' evaluation work or requested information about it if it was not publicly available.

4.3. Data analysis

The data from the interviews and workshop were summarised. Different statements were coded and clustered to gain deeper insights into the three main topics in the interview guide. We summarised and coded the findings from each of the document reviews. We also reviewed written descriptions of the organisations' evaluation schemes and activities. We registered any evaluation activities under the auspices of the public organisations (scope, frequency and focus of such evaluations) and any documented use of the findings. Furthermore, we reviewed project appraisal and planning documents related to later projects. When reviewing the documents, we systematically registered whether they referred to previous evaluations (i.e., by Concept, the organisation itself, or others).

Finally, we analysed all the data in a qualitative inductive manner by coding different statements and other pieces of information. Several themes had been identified by the researchers before the analysis and were used to guide it: the three steps of knowledge, use and benefit of evaluations; the relative emphasis of the evaluation model's various parts (e.g. performance of the project itself, as measured by 'efficiency', versus tactical and strategic measures of project success); the use of Concept evaluations (external) versus the organisations' evaluations (internal); and the causes of lack of use as discussed in the extant literature. We also looked for new, emerging themes. We were particularly interested in examples of good practices that could improve the knowledge, use, and benefit of evaluations.

Using multiple observers, data from various sources and combined methods are all effective means of triangulation, which enables a broad perspective and is crucial to strengthening a study's trustworthiness [43]. Moreover, including several organisations and sectors enabled researchers to note similarities or convergence of information that could strengthen the study's credibility.

5. Results

This section presents the findings from the interviews, workshop and document study. In line with the structure of our interview guide, we first look at public organisations' awareness of Concept's evaluations of their projects. Second, we look at their use of the evaluations (given that there is awareness). Third, we consider the ex-post evaluations carried out by the actors themselves and the use of these. Finally, we discuss key improvement points to increase general knowledge and the use and benefits of evaluations.

5.1. Awareness of evaluations and their results

5.1.1. Limited knowledge of evaluations. There was great variation among the interviewees regarding their knowledge of the evaluations. Some did not know the evaluations, some had become acquainted with them in preparation for the interview, some had been involved as informants in specific evaluations, and some had been made aware of completed evaluations through the programme's newsletter or website. Most interviewees were unsure of how well-known the evaluations were in the rest of their organisation. Still, their general impression was that the knowledge could be better.

However, the interview process helped increase knowledge about the evaluation scheme. For example, one interviewee said, "I read through one of the evaluations when preparing for the interview, and it was very interesting." Several interviewees mentioned that others in their organisation might be interested in the evaluations but probably did not know about them.

5.1.2. The evaluation model appears to cover what the target groups consider important. Although the interviewees' knowledge of the evaluations was limited, most expressed that the model appeared to cover what they considered important to evaluate and that the broad approach with six evaluation criteria was fitting for that purpose. However, there is a distinction between what is important for the project owner and decision-maker to evaluate and what is important for those who carry out the project. While the agencies are most concerned with what is going on in the delivery of the project (operational success), the ministries are more concerned with the wider societal effects of a measure (tactical and strategic success).

The ministries and agencies must uncover the causal relationship behind a good or bad result, not just the actual results. A separate chapter on "causal explanations and lessons learned" has been included in the most recent evaluation reports.

5.1.3. Learning from past projects is considered important. All interviewees were concerned with learning from and drawing on previous experiences when planning new projects. Some expressed that they learned in other ways than through evaluations, but many acknowledged that evaluations are also a crucial learning source. One interviewee said, "It can probably be as [important] to learn from a project that has gone particularly well as [from] a project that has gone badly."

Several interviewees pointed out they acquired knowledge through experience seminars, reference projects, collaboration, and a good organisational learning culture. The findings from the document study showed that most of the examined actors produce final reports for all completed projects, whereby information is stored so that it is available to the entire organisation. In some cases, the final reporting is routine, and it is uncertain whether the organisation will use the information well in subsequent projects.

Many of the interviewees emphasised the importance of using experience-related data in the planning of new projects. This focus was especially on the cost and schedule side; it was not on benefits. The extent to which the follow-up of benefits was practised systematically in the organisations varied. In some cases, poor information flow within and between adjacent organisations prevented them from using the experience gained in the past.

5.2. Use of evaluations in the planning of future projects

From the previous sub-section, we conclude that although ex-post evaluation is considered important, and the Concept evaluation model is considered relevant, there is too little knowledge of the evaluation scheme in the target group.

The natural consequence of limited knowledge is limited use (and therefore limited benefits). Our review of project appraisal and planning documents showed that they rarely refer to previous evaluations. (Neither Concept evaluations nor any other evaluations). Our informants pointed to some possible explanations for the limited use, in addition to the lack of awareness:

5.2.1. The evaluation of project delivery comes too late for the agency to make good use of it. The timing of an evaluation can significantly impact whether the results are perceived as relevant. Most of the agency interviewees stated that they were particularly interested in the performance targets for time, cost and quality (i.e., the efficiency criterion). In principle, it is possible to evaluate efficiency as soon as a project is completed, and the final cost is available. However, to address the other five criteria in the evaluation model, it is necessary to wait approximately five years into the operational phase before evaluating. This explains why Concept conducts its evaluations +/- 5 years after completion.

5.2.2. Limited learning potential from old projects. If we follow the above, evaluations must not be too old to be useful for the agencies. According to one interviewee, "Project assessments that go far back in time can have a somewhat limited learning effect today, as much has changed since the project appraisal was completed, and much of the information may be outdated". Additionally, the agencies' representatives pointed out that they worked differently than when the evaluated projects were initiated. Their suggested solution would be to conduct the evaluations earlier. However, there is a dilemma, as discussed below:

5.2.3. Evaluation of tactical and strategic project success requires that some time has passed. From the ministries' perspective, it is important to obtain information about the greater effects of a project and not just whether the delivery of the project has been successful. Normally, it takes several years from when a project is completed until the effects at the user and societal level become known, and hence, when it is appropriate to evaluate that project.

Some of the interviewees from ministries suggested that the agencies should be more concerned with project impact and societal goals. Even though the ministries own the societal goals, the choices made during the project delivery can often affect whether achieving the desired impact and societal goals in the long term is possible. One interviewee suggested that the evaluation process should be subdivided such that the efficiency criterion would be evaluated shortly after project completion, and the other five criteria would be evaluated later.

5.3. In-house evaluation

5.3.1. Few, if any, evaluations are carried out in-house by the ministry or agency. We found little or no ex-post evaluation of projects in the examined organisations. Apart from the Public Roads Administration, which conducts ex-post cost-benefit analyses [33], there is no systematic ex-post evaluation of projects in the examined agencies and no focus on evaluating the societal goals and impacts of projects. Although a few evaluations had been carried out in-house, they were typically triggered when there was a problem in the delivery of a project. These evaluations had more of a control function rather than learning as their focus.

The parent ministry must require an agency to carry out an ex-post evaluation. Without such requirements, the interviewed representatives of agencies considered it outside their mandate to evaluate the societal goals and impacts of projects. However, unclear ownership of the project's goals was a recurring theme that needed clearer accountability and follow-up. In several cases, little or no attention had been paid to societal goals after a decision had been made to implement a project.

5.4. Prerequisites for successful evaluation

5.4.1. The importance of finding the right contacts in the evaluated organisations. Before an evaluation, researchers from Concept always seek endorsement from the ministry responsible for the project. While some ministries had shown great interest in the evaluations and had even made suggestions regarding other projects they wanted to have evaluated, other ministries were uninterested or expressed that they did not see the need for an evaluation to take place. The interviewees suggested that this might depend on who they had encountered. The evaluation team is often quickly referred to a contact person; in practice, much of the subsequent contact will go through that person. During the interviews, the interviewees suggested that broader anchoring should occur at the beginning of the evaluation; one interviewee said, "We did not realise that this evaluation was underway. A start-up meeting with us or a round of interviews in the ministry would have been useful."

5.4.2. A need for more involvement in the evaluation process. A contact person from the ministry and the agency is appointed for every evaluation. Still, the contact person's degree of involvement in the design of the evaluations and during the evaluation process itself has varied. One explanation for the contact person's lack of involvement might be that Concept has left much of the contact with the target group to external consultants who have evaluated on the research programme's behalf: "The consultant was interested in us as interviewees, but not as a target group for the results. They should have involved us more in the design of the evaluation scheme itself". Several other interviewees shared this view. In an evaluation, the role of an interviewee differs from that of the user of that evaluation. The evaluator can easily forget the latter if they do not show sufficient interest in allowing the ministry and agency to influence the evaluation design and ensure the relevance of the evaluation. One of our interviewees exemplified this argument: "There was a clear plan for the evaluation when we became involved, and we were not asked to provide input into the scheme."

To date, all persons who have contributed as participants in an evaluation conducted by the Concept Research Programme have had the opportunity to comment on a draft report to ensure that the facts presented in the evaluation correspond to reality. However, some interviewees who had contributed in that way felt that their input had not been taken into account, and one interviewee who had commented on a draft said, "I was left with the impression that the comments we made were not considered sufficiently. This weakened our ownership of the evaluation results."

Some interviewees wanted to be more involved in discussions of the results and their implications, and several pointed out the need for more involvement. One of the interviewees suggested that there should be three regular meetings: "There should be (1) a meeting at start-up, (2) a meeting during the evaluation, where one can discuss preliminary findings and their implications, and (3) a meeting at the end where one is presented with the results." Another suggestion was that an evaluated party contribute to the evaluation report's chapter on "Causal explanations and lessons learned" in collaboration with the evaluator.

When the interviewees were asked whether they would have liked to participate with members of the evaluation team, either in part of an evaluation or an entire evaluation, several of the agency representatives stated that it could interest them. The involvement of someone from the evaluated organisations (although without a direct connection to the project being evaluated) could shorten the path from evaluation to use of the results without necessarily being at the expense of the evaluation's objectivity. Testing such involvement would be a matter of the availability of time and resources.

6. Discussion

In this paper, we have presented the findings from a study using external ex-post evaluations of large government projects from four sectors: roads and railways, buildings, ICT, and defence. The context is one country, Norway, where project evaluations are carried out under the auspices of a research programme funded by the Ministry of Finance and follow a standardised, six-criteria evaluation framework. The main purpose of these evaluations is learning and improvement in subsequent projects.

The reported experiences differ from the experiences of many other countries where evaluations of public projects are not carried out. Our findings demonstrate that neither lack of evaluation nor standardisation is a problem. Furthermore, the findings indicate that the poor quality of the reports, another cause of 'lack of use' identified in the literature, can be ruled out as the evaluations have been comprehensive and managed by experienced researchers. Moreover, Norway has extensive ex-ante evaluation processes, which provide excellent opportunities for experiences gained from previous projects to be used in planning new projects. Thus, the country should be well placed to ensure project evaluations are used.

Despite the good conditions, we find that the knowledge, use and benefits of these evaluations have been limited. Two main reasons for this finding are (1) lack of anchoring and participation, a common problem when an external party carries out evaluations, and (2) lack of integration into the organisations' project governance schemes. Our study found that public organisations do not evaluate their projects—all the more reason for them to make the most out of these evaluations conducted by external parties.

Many interviewed employees from ministries and agencies were unaware of the evaluations. Therefore, for the most part, the results had not been used for organisational improvements and to increase the quality of the planning, appraisal, and implementation of new projects. The evaluations have not been sufficiently well anchored in the responsible organisations, and the results of completed evaluations have not been disseminated sufficiently widely. Oxera [35] reported similar findings, namely that many stakeholders were unaware of previous evaluations or their findings.

It should be noted that our interviewees considered the broad, multicriteria evaluation framework suitable for assessing project success ex-post. Thus, lack of relevance does not seem to be a problem. However, we found indications that relevance could be increased by optimising the timing of an evaluation and by adapting the presentation of findings to the various target groups.

Furthermore, improved use could be achieved through early buy-in in the evaluations from the responsible organisations. This finding coincides with findings made by Johnson et al. [36], Fleischer and Christie [26], and Milzow et al. [38], all of whom have highlighted stakeholder involvement as crucial for the use of evaluation results.

We suggest that future evaluations emphasise target group involvement, not as individuals but as true representatives of their organisations with a clear mandate. This might improve the evaluations, as early input can improve evaluation designs. There is a risk that the external evaluations, initiated and carried out by researchers, are regarded as isolated activities outside the formal obligations of an organisation. To ensure a better link between formalised ex-ante appraisals and ex-post evaluations, appraisals should be required to refer to relevant reference projects and, if possible, to evaluations.

Knowledge management in project-based organisations is demanding, and individuals within those organisations first have significant experiences and, through them, develop their knowledge. Organisations depend on such individuals to improve their organisational skills. While there may be a shortage of good systems for communication within organisations, individuals learn in different ways, and it is probably unreasonable to require many people to read evaluation reports that are 80–100 pages long. Therefore, short and informative summaries and involvement in the evaluation process are important. Nevertheless, those responsible for all evaluations should have a dissemination strategy with different channels and mediums for communication with different target groups and individuals.

Regardless of the above suggestions for improvements, the needs of the responsible organisations should be at the forefront of any evaluation. The findings presented in this paper suggest that other issues may be more relevant for improving future appraisals and decision-making. There may be missed opportunities for learning from the evaluations we have investigated, but there may also be missed opportunities through different evaluation designs. Thus, evaluating the evaluations and their use should be a permanent activity.

7. Conclusion

Making evaluations relevant and useful for the target group without compromising the independent role of the evaluator will always be a dilemma. However, this dilemma is less relevant when the main purpose of evaluation is learning instead of control. Nevertheless, even when evaluations are relevant and results standardised, there are barriers to their use. This paper has suggested some improvement measures. Our study can contribute to improved practices and serve as a database from which ministries and agencies can learn when planning new investment projects. While our findings are not readily generalisable to other countries, they should be of interest to public organisations responsible for overseeing public investment projects.

References

[1] Hodgson D, Fred M, Bailey S, Hall P, editors. The projectification of the public sector. London: Routledge; 2019.
[2] Flyvbjerg B, Gardner D. How big things get done: The surprising factors behind every successful project, from home renovations to space exploration. London: Macmillan; 2023.
[3] Jong G de, Vignetti S, Pancotti C. Ex-post evaluation of major infrastructure projects. Transp Res Procedia. 2019; 42:75–84.
[4] Flyvbjerg B. Quality control and due diligence in project management: Getting decisions right by taking the outside view. Int J Project Manage. 2013;31(5):760–74.
[5] Morris P. Reconstructing project management reprised: A knowledge perspective. Proj Manage J. 2013; 44(5):6–23.
[6] Williams T, Samset K. Issues in front-end decision making on projects. Proj Manage J. 2010; 41(2):38–49.
[7] Williams T, Klakegg OJ, Walker DHT, Andersen B, Magnussen OM. Identifying and acting on early warning signs in complex projects. Proj Manage J. 2012;43(2):37–53.
[8] Williams T, Vo H, Samset K, Edkins A. The front-end of projects: a systematic literature review and structuring. Prod Plan Control. 2019;30(14):1137–69.
[9] Samset, K, Christensen, T. Ex-ante project evaluation and the complexity of early decision-making. Public Organization Review. 2017; 17, 1-17.
[10] Worsley, T. Ex-post Assessment of Transport Investments and Policy Interventions: Prerequisites for Ex-post Assessments and Methodological Challenges. International Transport Forum Discussion Paper No. 2014-19; 2015.
[11] Mackie P, Preston J. Twenty-one sources of error and bias in transport project appraisal. Transp Policy. 1998;5(1):1–7
[12] Nicolaisen MS, Driscoll PA. An international review of ex-post project evaluation schemes in the transport sector. J Environ Assess Policy Manag. 2016;18(01):1650008
[13] Flyvbjerg B, Bruzelius N, Rothengatter W. Megaprojects and risk: An anatomy of ambition. Cambridge: Cambridge University Press; 2014
[14] Volden, GH and Samset, K. Quality assurance in megaproject management: The Norwegian way. Flyvbjerg, B. (ed), *The Oxford Handbook of Megaproject Management*. Oxford: Oxford University Press; 2017.
[15] Finansdepartementet. Reglement for økonomistyring i staten: Bestemmelser om økonomistyring i staten; 2021.
[16] OECD DAC. DAC Principles for Evaluation of Development Assistance; 2019. Available from: https://www.oecd.org/dac/evaluation/2755284.pdf
[17] Volden GH. Public project success as seen in a broad perspective. Eval Program Plann. 2018; 69:109–17
[18] Welde M, Volden GH. Measuring efficiency and effectiveness through ex-post evaluation: Case studies of Norwegian transport projects. Res Transp Bus Manag. 2018;28:33–9
[19] Dahler-Larsen P. The Evaluation Society. Palo Alto, CA, USA: Stanford University Press; 2013

IOP Conf. Series: Earth and Environmental Science **1389** (2024) 012025 doi:10.1088/1755-1315/1389/1/012025

[20] Samset K. Reell eller rituell evaluering? Stat Styr. 2015;25(3):6–9
[21] Alkin MC, Daillak RH. A study of evaluation utilization. Educ Eval Policy Anal. 1979;1(4):41.
[22] Alkin MC. Using evaluations: Does evaluation make a difference? Thousand Oaks, CA, USA: SAGE Publications; 1979.
[23] Knorr, KD. Policymakers' use of social science knowledge: Symbolic or instrumental? Weiss, C.H. (ed), Using Social Research in Public Policy Making. Lexington, MA: Lexington Books; 1977.
[24] Weiss, CH. Utilization of evaluation: Toward comparative study. Weiss, C.H. (ed), Evaluating Action Programs: Readings in Social Action and Education. Boston, MA: Allyn and Bacon; 1972.
[25] Henry GT, Mark MM. Beyond use: Understanding evaluation's influence on attitudes and actions. Am J Eval. 2003;24(3):293–314.
[26] Christie CA. Evaluation use: Results from a survey of U.S. American evaluation association members. American Journal of Evaluation. 2009;30:158–75.
[27] Dahler-Larsen, P. The sceptical turn in evaluation. Furubo, J.E. and Stame, N. (eds), The Evaluation Enterprise. New York: Routledge; 2018
[28] Vedung, E. Six uses of evaluation. Hennefeld, V, Meyer, W and Silvestrini, S (eds), Nachhaltige Evaluation? Auftragsforschung Zwischen Praxis und Wissenschaft. Festschrift zum 60. Geburtstag von Reinhard Stockmann, Waxmann, Münster; 2015.
[29] Eliasson J, Börjesson M, Welde M. Does benefit-cost efficiency influence transport investment decisions. Journal of Transport Economics and Policy. 2015;49:377–96.
[30] Dahler-Larsen, P. Surviving the routinization of evaluation: The administrative use of evaluations in Danish municipalities. Administration and Society. 2000; 32, 70-92.
[31] Agenda Kaupang and Rambøll Management Consulting. Bruk av evaluering i statlig styring. Direktoratet for økonomistyring rapport R9392; 2016.
[32] Meunier, D, Wortelboer-van Donselaar, PW and Fitzroy, S. Ex-post analysis and tools: What should you expect? Paper presented at the 44th European Transport Conference, Barcelona, 5–7 October; 2016.
[33] Odeck J, Kjerkreit A. The accuracy of benefit-cost analyses (BCAs) in transportation: An ex-post evaluation of road projects. Transp Res Part A Policy Pract. 2019;120:277–94.
[34] Ledermann S. Exploring the necessary conditions for evaluation use in program change. Am J Eval. 2012;33(2):159–78
[35] Oxera. How should the ex-post evaluation of trunk road schemes be enhanced? Report prepared for Department for Transport; 2005.
[36] Johnson K, Greenseid LO, Toal SA, King JA, Lawrenz F, Volkov B. Research on evaluation use: A review of the empirical literature from 1986 to 2005. Am J Eval. 2009;30(3):377–410.
[37] Berriet-Solliec, M, Labarthe, P and Laurent, C. Goals of evaluation and types of evidence, Evaluation. 2014; 20, 195-213.
[38] Milzow K, Reinhardt A, Söderberg S, Zinöcker K. Understanding the use and usability of research evaluation studies. Res Eval. 2019;28(1):94–107.
[39] Direktoratet for økonomistyring. Veileder: Strategisk og systematisk bruk av evaluering i styring; 2011.
[40] Bastøe PØ. Evalueringsvirksomheten må institusjonaliseresfor å heve kvaliteten. Stat Styr. 2020;30(2):53–5.
[41] Scriven, M. Key Evaluation Checklist (KEC); 2013. Available from: http://michaelscriven.info/images/KEC_7.25.2013.pdf
[42] Bukkestein, I, Welde, M and Volden, GH. Bruk og nytte av etterevalueringer av prosjekter. Concept arbeidsrapport 2020-3, Norwegian University of Science and Technology, Trondheim; 2020.
[43] Lincoln YS, Guba EG. Naturalistic Inquiry. Thousand Oaks, CA, USA: SAGE Publications; 1985.