

Abstract

This article analyses the Norwegian government's evaluation practice over the 25-year period from 1994 to 2018. Evaluations are mandatory for government ministries and agencies in Norway, with the government conducting some 100 evaluations annually. This article utilises data from a unique database to describe the development of the evaluation industry, focusing on the volume of evaluations, the most active commissioners and providers of evaluations, and the types of evaluations conducted. First, the analysis indicates that the volume of evaluations peaked in around 2010 and has subsequently decreased. As a possible consequence, information relevant to policy may be less publicly available than before. Second, ministries have commissioned relatively fewer evaluations in the last decade than in the years before, and executive agencies have commissioned relatively more. Third, the proportion of evaluations performed by consultants has risen, with that of research institutes falling.

Introduction

In his keynote speech at the launch of the Norwegian Evaluation Society in Oslo on 27 May 2009, Murray Saunders, President of the European Evaluation Society, stated that “evaluation practice, as an object of research, is completely underdeveloped”. Since then, a growing body of research on evaluation has emerged (Christie and Alkin 2013; Coryn et al. 2017; Nielsen et al. 2018), thereby advancing systematic knowledge about the methods used in evaluations (Kallemeyn et al. 2015; Lemire et al. 2020), the utilisation of evaluations in policymaking (Dahler-Larsen 2000; Leeuw 2009a) and prescriptive contributions on how to evaluate (Howell and Yemane 2006; Nutley et al. 2007). Despite advances, much remains to be done in analysing policy evaluation as an industry and a field of practice (Christie and Fleischer 2010; Nielsen et al. 2018). Research on evaluation practice has relied predominantly on examinations of selected evaluations (Howell and Yemane 2006) and expert surveys (Furubo et al. 2002; Jacob et al. 2015), as well as surveying reviews of existing research, not data on actual evaluation practice (Nielsen et al. 2018: 15; Coryn et al. 2017).

***Jostein Askim** is a professor in the Department of Political Science, University of Oslo, Norway and a professor II at the Department of Administration and Organization Theory, University of Bergen, Norway. He studies public sector reforms, public management and political leadership, with a particular focus on the Nordic countries. His research has appeared in, for example, *Journal of Public Administration Research and Theory*, *Public Administration*, *Government and Opposition*, and *Political Studies Review*.

Erik Døving is an associate professor in organization science at Oslo Business School, Oslo Metropolitan University. His research interests include leadership, organizational learning, evaluation studies and human resource management. His work has been published in for example *Strategic Management Journal*, *International Small Business Journal*, *Journal of Workplace Learning*, *Research Studies in Music Education*, and *Leadership*.

Åge Johnsen is Professor of Public Policy at Oslo Business School, Oslo Metropolitan University. He is also Adjunct Professor at the Defence Command and Staff College, The Norwegian Defence University College. His research interests are public sector reforms, strategic planning, performance management, trust-based management, change management, and performance audit and evaluation in the public sector. His research has been published in, for example, *Evaluation*, *International Review of Administrative Sciences*, *Public Management Review*, and *Financial Accountability and Management*.

Jostein Askim,
Department of Political
Science, University of Oslo,
Norway
jostein.askim@stv.uio.no

Erik Døving,
Oslo Business School, Oslo
Metropolitan University,
Norway
erik.doving@oslomet.no

Åge Johnsen,
Oslo Business School, Oslo
Metropolitan University,
Norway
aage.johnsen@oslomet.no

Keywords:
research on evaluation,
government evaluations,
evaluation industry,
Norway

In this article, we analyse government evaluation practice as it has developed over time in one country: Norway. As far as we know, the study is the first of its kind to be based on large-N systematic and longitudinal data. We study the government's evaluation practice using data from the Evaluation Portal, which is the Norwegian Agency for Public and Financial Management's vehicle for collecting, archiving and publicising reports from government-commissioned evaluation projects. Our aim is to study evaluation practices as revealed through actual evaluations undertaken by organisations external to the government agency responsible for the policy that is being evaluated. Our analysis covers more than 3,000 evaluation reports produced by organisations including research institutes, consultancy firms, ad hoc committees, and the Office of the Auditor General. Norway's evaluation practice has escaped scholarly attention over the past 20 years, apart from a handful of studies covering the 1980s and 1990s (Eriksen 1990; Baklien 1993; Brofoss 1997; Øvreliid 2000; Wærness and Øvreliid 2003). Furthermore, Norway's evaluation practice has not been analysed using systematic data, with the exception of Askim et al.'s (2013) study focusing on the period 2005–2011. This article significantly extends the period studied.

We examine how the volume of evaluations registered has developed over the 25-year period between 1994 and 2018 and, on this basis, discuss whether, as some have claimed, evaluation appears to be an ever-expanding activity. We also examine whether evaluation is primarily oriented towards policy interventions that have been completed (summative evaluation), towards ongoing policy interventions (formative evaluation) or towards policy interventions at the planning stage (prospective evaluation). In addition, we look at how Norway's evaluation industry has developed over the past 25 years. By evaluation industry, we mean the commissioners of evaluations, evaluators and their respective relations with the policy owners. These issues are related to the origin of evaluations (Henry and Mark 2003) and are central to widely cited assessments and comparisons of national evaluation practices (Furubo et al. 2002; Jacob et al. 2015). We also investigate what changes have occurred in the types of government entities that commission evaluations. For example, has the appetite for evaluations spread to all policy areas, and which changes have occurred in relation to evaluators; and have academic organisations such as universities and research institutes become more or less central in the field over time, compared with evaluators in the consultancy industry?

The remainder of this article is structured as follows. In the next section, we define and delineate the evaluation concept and the evaluation industry. In the third section, we present the research context and describe our data sources. We then describe the Norwegian context with its evaluation database and our analytical procedures. The subsequent sections report and discuss the results from the analyses, with the final section concluding the article with principal findings and recommendations for further research into evaluation practices.

The Industry: Evaluations, Commissioners and Evaluators

The Volume of Evaluations

Observers of evaluations, regardless of their more specific analytical perspective, seem to agree that the volume of evaluations is constantly increasing (Leeuw 2009a; Nielsen et al. 2018; Lemire et al. 2018b). Leeuw (2009b) notes that evaluation activities “blossomed” in the Netherlands post-2000. When the word waves is used to describe evaluation practice, it is not used as a metaphor for changes in volumes over time, as in “wave-shaped development”. Rather, the notion of waves is used to describe approaches to evaluation that come and go. According to Vedung (2010), since the 1960s, the public sector has experienced scientific, dialogue-oriented, liberal and evidence-based waves of evaluation. Scholars who advocate evidence-based governance invariably assume that evaluation is a growing phenomenon. By evidence-based, we are referring to the norm that policymaking ought to be based on systematic knowledge, not ideology and special interests (Pawson 2006; Sanderson 2002; Pfeffer and Sutton 2006). How volumes have developed over time is, nonetheless, an empirical question – one that we will examine based on the Norwegian case. If the volume of evaluations displays a pattern other than a steady increase over time, the notion that evaluation is an ever-increasing process should be revisited and revised.

Early studies of Norway’s evaluation practice found evaluation to be particularly frequent in certain policy areas, notably development aid, welfare and social policy, as well as education and research (Baklien 1993; Brofoss 1997). Similar patterns, with evaluation practice unevenly distributed across policy areas, have been found in Denmark (Hansen 2003), Sweden (Furubo et al. 2002; Furubo 2007), USA (Lemire et al. 2018a; Weiss 1987), and the Netherlands. Leeuw and Rozendal (1994) found that the Dutch central government commissioned around 200 evaluations annually. In the 1990s, the Danish government commissioned close to 100 evaluations per year (Hansen 2003). Balthasar (2009) found that the Swiss federal government commissioned some 80 evaluations per year in 1999 and 2000. In Norway, the government commissioned about 100 evaluations per year during the period 2005–2011 (Askim et al. 2013).

Types of Evaluations

Interest in the development of evaluation practice over time should not be limited to volume. It is also important to keep an eye on what types of evaluations governments commission, as this, in turn, affects what kind of information evaluations feed into political processes that are oriented towards policymaking and accountability. We do not focus on methodology, which is a variable sometimes used to distinguish between types of evaluation (Christie and Alkin 2013; Christie and Fleischer 2010). Instead, we focus on the timing of evaluations vis-à-vis their objects.

Summative evaluations are retrospective, i.e. they assess the value of a completed policy intervention or project, or of reforms that have been implemented. For example, summative evaluations assess whether policy interventions achieve their policy objectives, whether their positive benefits

outweigh the resources committed to them, and whether they were implemented cost-effectively. The intended audiences for summative evaluations are typically policy owners (e.g. government ministries) and the general public. Summative evaluations are usually commissioned from evaluators independent of the policy owner to avoid suspicion that evaluators have been pressured into producing a positive assessment (Vedung 2006).

Summative evaluations are often regarded as the core of evaluation (Bovens et al. 2006). Vedung, for example, defines evaluation as “retrospective assessment of the administration, output and outcome of public interventions (...) intended to play a role in future practical situations” (2005: 297). However, a narrow conception of evaluation means that a considerable proportion of knowledge production that is relevant to policy, and that is regarded as evaluation by practitioners, falls outside our scope. In this article, we also are interested in formative and prospective evaluations.

Formative evaluations assess time-bound, but ongoing reforms, policies and interventions. They can also assess permanent programmes and the performance of government organisations, with permanent meaning that there is no scheduled termination (e.g. a pension scheme). In some contexts, process evaluation is used synonymously with formative evaluation, but we use the latter. This type of evaluation is sometimes commissioned because policy owners want to have a knowledge base for adjusting ongoing operations. Although perhaps not as important as in the case of summative evaluations, commissioning evaluations from organisations that are independent of the policy owner can be a priority for formative evaluations as well. For example, in-house evaluators might be reluctant to criticise a project owned by their superiors and implemented by their colleagues, and can be blind to shortcomings in the policy intervention due to groupthink. Formative evaluations are often commissioned by the organisation that is implementing the policy intervention in question, a government agency for example, and not by the ultimate policy owner, for example, a government ministry. Evaluators also often report the results from formative evaluations directly to those executing the policy, rather than the policy owners, thus enabling a quick turnaround time from feedback to improvements to ongoing operations (Vedung 2006).

Finally, prospective evaluations assess proposed policy interventions based on theory or based on experience of similar policy interventions (General Accounting Office 1999; Scriven 1991). Economic cost-benefit analysis is one approach to prospective evaluation (Vining and Weimer 2006), but a prospective evaluation can also assess proposals without articulating their expected benefits and costs in economic terms.

Commissioners

Evaluations often criticise a policy and might initiate consequences for policy owners, including loss of popular support for political parties and office holders who championed the policy. The institutional distance between the policy owner, the actor who commissioned the evaluator (the commissioner) and the evaluator is therefore a topic of interest to many scholars in the field of evaluation (Vedung 2005, 2006; Balthasar 2009). We will start with the commissioners' position before discussing that of evaluators.

Policy owners sometimes commission evaluations themselves. For example, a government ministry that has implemented a reform can commission an evaluation of the reform. On other occasions there is an institutional distance between the ultimate policy owner – in many instances, a government ministry – and the commissioner. The ministry may not implement the policy or intervention in question itself and may want those who have day-to-day responsibility and operational expertise to manage the task of commissioning the evaluations and to constitute a buffer against any criticism that might emerge from the evaluation itself. Alternatively, the government might want to create a buffer between it and the commissioner because it wants to maximise the independence of the evaluation by ensuring that not only the evaluator, but also the commissioner, is independent. For example, governments sometimes ask organisations such as research councils to commission evaluations.

Evaluators

Across many countries, the maturing of the evaluation field has involved the development of a professional identity: establishment of professional organisations, academic study programmes and certifications, and development of an “evaluation industry” (Nielsen et al. 2018: 14). Evaluation has been called a growth industry (Leeuw 2009a) and has undoubtedly grown in most countries, in terms of the economic size of the market and the number of market operators vying for contracts, and in the sense that the marketplace comprises an increasingly heterogeneous set of operators. As Vedung (2010) describes it, different waves of evaluation have washed over the shores of the public sector, with each wave leaving a sediment (see also Leeuw and Furubo 2008). For example, today’s marketplace comprises everything from universities and applied research institutes associated with the scientific wave of the 1960s, to management consultancy firms associated with the liberal wave and the emergence of the so-called audit society (Power 1997). The evaluation marketplace also comprises tiny, highly specialised operators and huge, generalised operators.

Obviously, evaluators do not just comprise organisations outside the public sector; the public sector also has its own permanent and temporary evaluation capacities. For example, many governments have internal audit offices, and it is thus not entirely uncommon for the policy owner and evaluator to be one and the same (Vedung 2006), with nothing but paper walls between them. In many countries, supreme audit institutions (SAIs) are very central evaluators of public policies and government organisations. Norway is no exception, as the Office of the Auditor General is, as we show below, Norway’s largest evaluator. SAIs enjoy a degree of independence from the executive government and they should thus be viewed as semi-internal. The same goes for committees that the government appoints to perform an evaluation. Such ad-hoc committees are an essential aspect of the policymaking process in several countries, Norway included, and they conduct, for example, high-profile prospective and summative evaluations. An example of the latter is that many European countries – e.g. Norway, Denmark, Sweden and France – have appointed semi-independent commissions to evaluate their respective governments’ handling of the COVID-19 pandemic crisis.

Research Context

Norway is a unitary state with a population of about 5 million and a parliamentary system of government. The central government is organised into 16–20 government ministries (with some variations between governments) and about 150 national executive agencies (e.g. directorates and regulators). Developing and evaluating public policy is primarily the responsibility of central government. Much of the responsibility for implementing public policies is delegated to subnational government agencies and to the local government sector, comprising 10 counties and 356 municipalities. Note that any evaluation commissioned by a public sector organisation other than the central government falls outside the scope of this study.

In the last three decades, delegation of functions from ministries to agencies has left the ministries themselves at arm's length from much of the actual implementation of national policies (Askim et al. 2019). This trend has accentuated the general challenge of governing from a distance. Ministries are typically no longer directly involved with practical implementation, and there is a corresponding shortage of first-hand experience of the effects of public policy. Ministries consequently depend, now more than in earlier periods, on systematic monitoring of underlying agencies and their operations. In Norway, as in other empirical contexts, commissioning evaluations from external evaluators reflects this trend, as do requirements for internal audits (Andersen and Pedersen 2012).

Moreover, evaluating has become a mandatory activity for the Norwegian central administration. The Rules for Financial Management (Reglement for økonomistyring) require organisations in central government to commission retrospective evaluations, and prospective evaluation is a requirement in public management instructions as well as in the Public Administration Act. Likewise, contracts for evaluations are generally awarded after public tender – as per Norway's Public Procurement Act and European Economic Agreement regulations. Minor evaluations may be commissioned within existing framework contracts.

The Norwegian market for evaluation is dominated by consulting firms and research institutes (as well as the Auditor General, which mandates its own evaluations and does not compete for tenders in the open market). Norway has a large research industry comprised of not-for-profit research institutes. While formally independent (often organised as foundations), they enjoy semi-public status. The research institutes are rooted in an academic research culture and are typically staffed and managed like university departments. Research institutes are usually accredited by the Norwegian Research Council and are eligible for basic funding. Accreditation criteria stipulate that research institutes must differ from consulting firms (Research Council of Norway 2018; 2019). In the 2010s, some research institutes merged with each other and/or with government-owned universities and university colleges. The consulting industry spans a wide range from a small number of large, well-known international consulting firms to small “consulting offices” made up of a handful of employed or independent consultants. Consulting firms are typical professional service providers and some of the larger firms offer several services (management consulting, auditing, accounting, legal services, economic analysis, engineering, etc.) under the same

brand. Large consulting firms tend to be tightly managed and provide services based on a more standardised approach than that of, for example, research institutions (Løwendahl, 2016; Von Nordenflycht, 2010).

Data and Methods

We analysed data from the period 1994–2018, 25 years in all. The principal data was obtained from the Norwegian Agency for Public and Financial Management’s Evaluation Portal, which was established in 2010 as the platform for registering, cataloguing and publicising evaluations commissioned by organisations in the Norwegian government administration, i.e. government ministries, government agencies and government-owned enterprises (Norwegian Agency for Public and Financial Management 2020a). The Agency for Public and Financial Management subsequently delegated collection and registration of evaluations to the National Library of Norway. In establishing the portal, the Agency for Public and Financial Management had the aim of facilitating public access to evaluations by archiving evaluation reports in one place and making them easily accessible, thereby increasing the use of these reports – and, possibly, their impact – while also enhancing transparency in the public sector.

The Evaluation Portal is designed to comprise all evaluations in central government, regardless of how the evaluations have been commissioned. Registration of evaluations in the Evaluation Portal is voluntary, but registration of documents including evaluations in the National Library is mandatory. The Agency for Public and Financial Management, and subsequently the National Library, has followed up the commissioners’ and providers’ registration of their evaluations in the Evaluation Portal. However, this monitoring of under-reporting has been less systematic since around 2016 than it was prior to that. The volume of data in the Evaluation Portal thus to some degree probably understates the factual volume of evaluations in recent years. We triangulated the data by supplementing the time series data from the Evaluation Portal with time series data for evaluations registered in the National Library catalogue and the university library catalogue, see Figure 1.

The Evaluation Portal covers evaluations carried out by external actors for the Norwegian central government administration. In this context, external actors include organisations outside the public agency responsible for the policy, programme or project being evaluated. While evaluators are often independent contract research institutes, evaluation projects may also be commissioned from state universities or institutes – that is, evaluations are also considered to be external if undertaken by a public agency independent of the policy owner. Evaluations commissioned from ad-hoc committees of experts external to the agency being evaluated are usually appointed by a ministry, they are not subject to competitive tender in the market. Finally, reports from the Office of the Auditor General are included in the database. The Auditor General is formally a subsidiary of the Parliament, but enjoys wide autonomy and is independent of all public agencies that implement government policy. As mentioned above, the Auditor General mandates its own evaluations. Evaluations that the central government’s administrative organisations produce internally themselves (e.g. self-evaluation reports) are not recorded, and neither are evaluations that are

exempt from public disclosure. In addition to the evaluation report, the portal contains key/summary information, e.g. the evaluation's title, year of publication, commissioners (oppdragsgivere), evaluators and types of evaluations (Norwegian Agency for Public and Financial Management 2020b).

As we will see in the next section, the number of evaluations registered annually in the Evaluation Portal was relatively low during the 1994–2004 period, when the Agency for Public and Financial Management purposely only registered the most important evaluations. The underlying volume of evaluations was therefore higher during that period than indicated by the data (source: personal communication with officials from the Agency for Public and Financial Management and the National Library), as was probably also the case after 2016. However, there are no indications that the data pertaining to evaluations prior to 2005 or after 2016 is biased with regard to types of evaluations, commissioners and evaluators.

Another factor concerning the accuracy of the data is that the Research Council of Norway, an agency formally under the Ministry of Education and Research, is recorded as the evaluator in almost 250 instances (see Figure 3). Typically, this indicates when a ministry or agency has delegated administration of an evaluation to the Research Council. In such cases, the Research Council of Norway appoints a committee of experts to manage the coordination of various parts of the evaluation. What is registered in the evaluation portal is typically a summary report written by the committee, not reports from the underlying sub-projects. For example, in 2001, the Research Council of Norway was registered as the evaluator of a family support scheme (kontantstøtteordningen). The report was authored by three academics who were members of the research council committee set up to coordinate the evaluation. The report summarises and quotes seven sub-evaluations. Reports from the sub-evaluations, each produced by a research institution, were not registered separately in the portal; only the summary research council report was registered. There are no indications that the practice of registering sub-evaluations varies systematically across government ministries or over time. However, we can assume that evaluations registered as being produced by the Research Council of Norway were also commissioned by the Research Council, and that when the Research Council of Norway is recorded as the evaluator, it represents an underreporting of research institutes as evaluators.

The evaluation database records evaluation reports as one or more of five types: cost-benefit analysis (samfunnsøkonomisk analyse); prospective evaluation (forhåndsevaluering); process evaluation; summative evaluation (etterevaluering); and time-independent evaluation. Examples of evaluations registered as time-independent are auditor general reports on government agencies' operations and systematic literature reviews of a field of scientific study. The database allows for recording more than one type per evaluation report, and each report included in this study had been recorded on average as 1.1 types. One reason for ticking more than one box is that the database's five categories are not mutually exclusive. As explained in the registration template, "cost-benefit analysis is a subcategory of prospective evaluation, so please register the report as both" (Norwegian Agency for Public and Financial Management (2020b)). Another reason for ticking more than one box is that

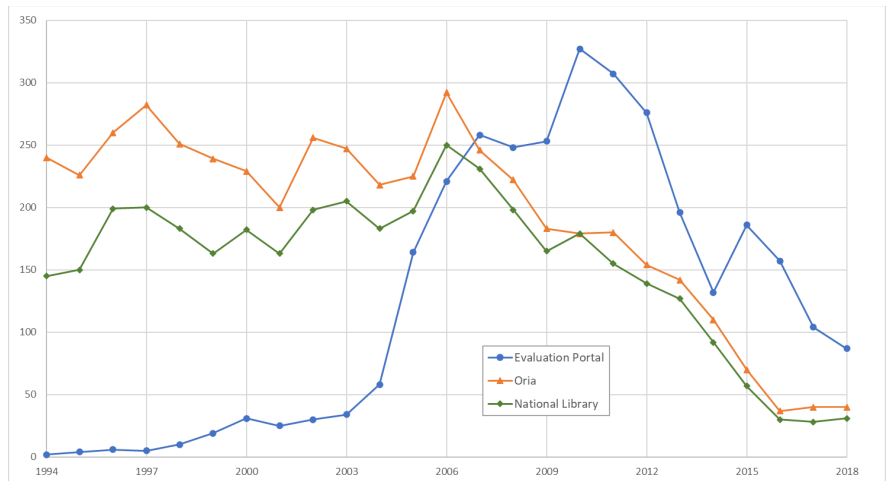
evaluation reports sometimes cover different aspects of a larger evaluation project, e.g. one policy intervention that is ongoing and one that has been concluded. In this study, we view evaluations registered as cost-benefit analyses as prospective evaluations, and evaluations registered as process and time-independent evaluations as formative evaluations (see Figure 2 below).

Evaluation Practices in Norwegian Central Government – 1994–2018

Volume of Evaluations

We start by analysing how many evaluations are actually performed in the Norwegian central administration. Figure 1 shows the number of evaluations per year between 1994 and 2018, using data from three sources. The pattern is clear: the volume of evaluations increased through the 2000s, peaked in 2010 and has decreased during the 2010–2020 period. Supplementary analyses of the university libraries’ joint catalogue and the catalogue of the National Library reveal a similar pattern, with a notable downward trend during the last five to ten years.

Figure 1. Number of evaluations in Norway per year, as registered in three databases, 1994-2018



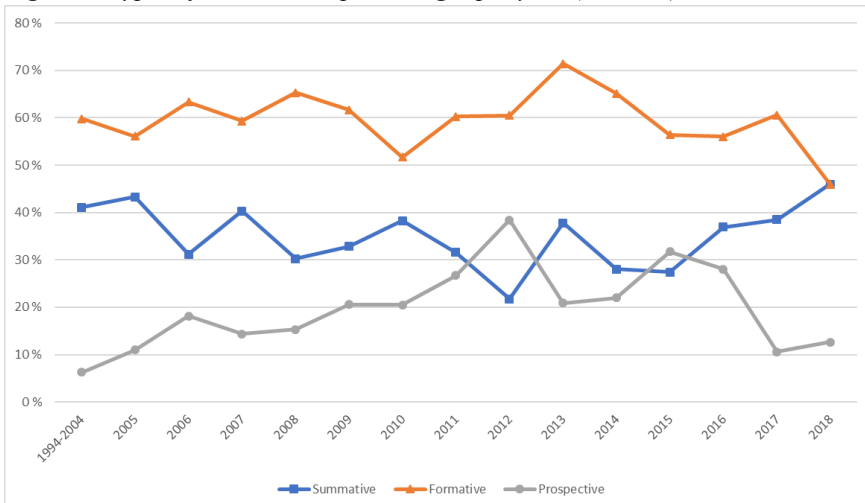
Sources: The Evaluation Portal. Oria is the joint catalogue for the libraries of Norwegian higher education and research institutions. The National Library catalogue covers, inter alia, all publications in Norway submitted in accordance with legal deposit. We searched the latter two databases for “evaluation of” (Norwegian: “evaluerende av”) limited to research literature (“faglitteratur”).

Types of Evaluations

In the conceptual section, we distinguished between prospective, formative and summative evaluations. Figure 2 shows the percentage of types of evaluations per year since 1994. Formative evaluations have generally been the most frequent type, at around 60 per cent of the total, though they dipped in frequency in 2018. However, we are reluctant to suggest that this can be attributed to anything other than a random fluctuation. Prospective evaluations challenged summative evaluations for second place in the early 2010s, but summative

evaluations have generally been the second most frequent type. Formative and summative evaluations were joint most frequent in 2018, each at about 45 per cent of the total volume of evaluations. Formative evaluations might be considered to be more internally oriented, suitable for helping government organisations to improve ongoing policies, while summative evaluations are more externally oriented, suitable for equipping external stakeholders and the citizenry with information to make judgements about the government’s ability to solve social problems. Based on the quantity of each of the two types of evaluations, we cannot point to any particular trend, for example, that government evaluation has become more internally or more externally focused over time. There appear to be more fluctuations over time in prospective evaluations than in formative and summative evaluations. In the past two years, only some 10 per cent of the evaluations have been of the prospective type, however, this can be a random fluctuation because prospective evaluations normally constitute more than 20 per cent of the total.

Figure 2. Types of evaluations, percentages per year (N=3140)



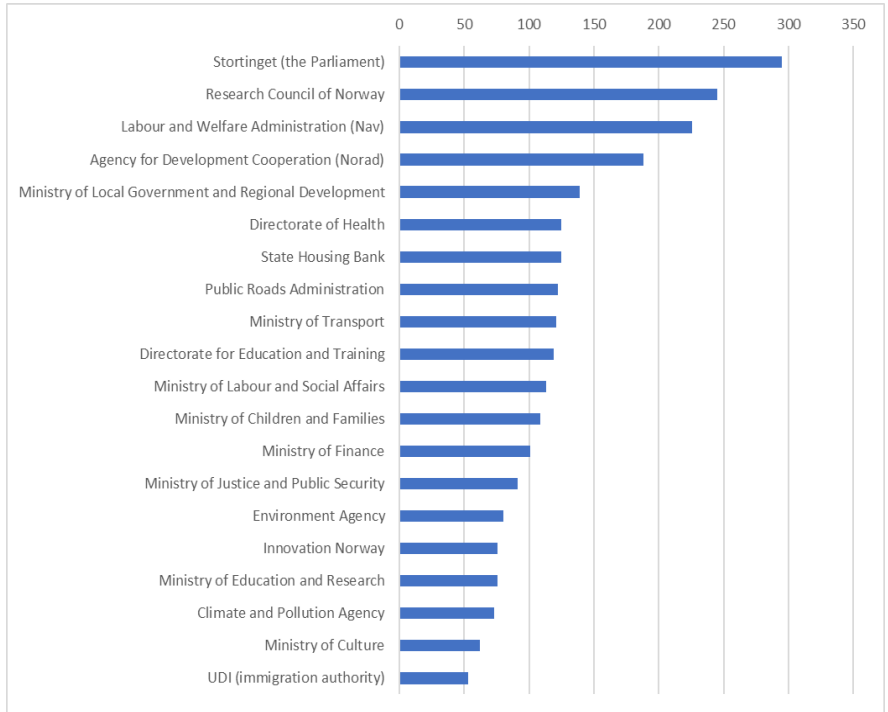
Source: Evalueringsportalen.no. Evaluation reports can be recorded as more than one type, so overall annual percentages total more than 100.

Prospective evaluations occur particularly often in the transport and energy policy sectors (predominantly cost-benefit analyses), while summative evaluations occur particularly often in the employment, welfare, housing and environmental policy sectors. Formative evaluations occur particularly often in the public administration policy sector, with many formative evaluations constituting reports from the Office of the Auditor General concerning a government agency or policy sector.

The Commissioners

Figure 3 shows the top 20 commissioners in Norway between 1994 and 2018, measured according to the number of evaluations. Taken together, these 20 organisations commissioned more than one third of all evaluations in the period.

Figure 3. Top 20 commissioners and numbers of evaluations commissioned by each, 1994–2018 (N=2539).



Source: Evalueringsportalen.no.

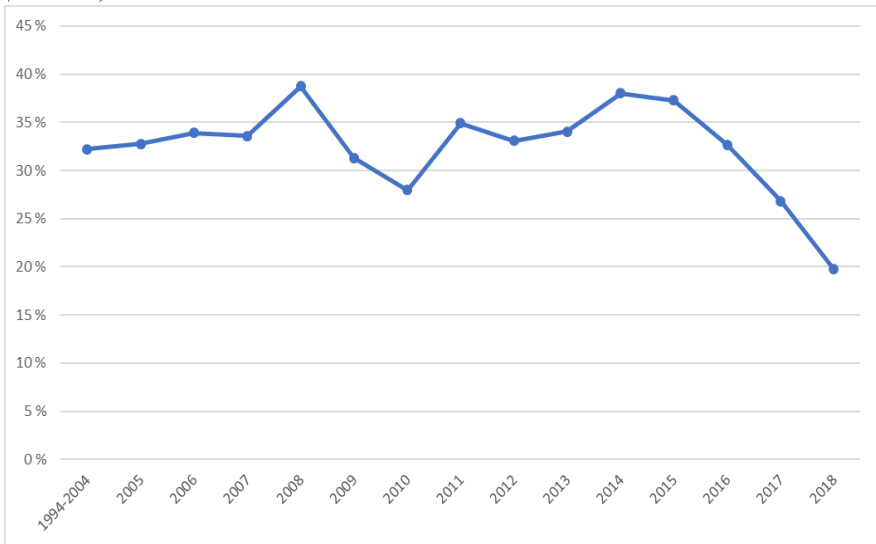
Topping the list is Parliament, with almost 300 evaluations, i.e. about 12 evaluations per year, most of which are evaluations conducted by the supreme audit institution, the Office of the Auditor General (OAG), which routinely submits financial audits of government ministry accounts and performance audits to Parliament. An example of a performance audit from the Office of the Auditor General is “Study of the digitalisation of the cultural heritage”, a report from 2017. Note that only rarely does Parliament ask the supreme audit institution to evaluate a specific entity, such as a specific reform or government organisation. Parliament can therefore not be viewed as an ordinary commissioner of evaluations in cases in which the Office of the Auditor General is the evaluator. Second on the list is the Research Council of Norway, also no ordinary commissioner of evaluations (see the methods section). One could therefore argue that the organisations that follow in the top 20 list are the most active commissioners: the Employment and Welfare Administration, the Agency for Development Cooperation, and the Ministry for Local Government and Regional Development.

The fact that ministries and agencies responsible for the employment and welfare, international aid and development, and transport and infrastructure policy sectors top the list confirms patterns found by researchers into Norwegian evaluation practice in the 1980s and early 1990s. However, the Ministry of Education and Research seems to have commissioned fewer evaluations in the past 25 years than in earlier periods (Baklien 1993; Brofoss 1997). The fact that

this ministry and the Directorate for Education and Training (established in 2004) are not at or near the top of the list of commissioners is also surprising in the light of findings from Denmark (Hansen 2003) and the Netherlands (Leeuw and Rozendal 1994), where education policy institutions are particularly active commissioners of evaluations.

The ministries that ordered the smallest number of evaluations in Norway between 1994 and 2018 were the Ministry of Fisheries and Coastal Affairs, the Ministry of Agriculture and Food, and the Ministry of Defence. Organisations concerned with agricultural policy also commissioned a small number of evaluations in Denmark (along with finance and agriculture; see Hansen 2003) and in the Netherlands (along with finance and justice; see Leeuw and Rozendal 1994).

Figure 4. Evaluations commissioned by government ministries as a proportion of all evaluations per year. Evaluations commissioned by Parliament excluded (N=3460).



Source: Evalueringsportalen.no.

Returning to the case of Norway, if we disregard evaluations commissioned by Parliament, government ministries commissioned about one third and executive agencies about two thirds of all evaluations between 1994 and 2018. Figure 4, which shows the proportion commissioned by ministries, illustrates the fact that the ministry/agency pattern was relatively stable for 20 years before the proportion commissioned by ministries declined markedly in the final years under study. The proportion in 2018 was 20 per cent, compared with almost 40 per cent in 2008, 2014 and 2015. If the recent trend continues, evaluation will become an agency-related practice, with ministries virtually ceasing to commission evaluations themselves. As an illustration, foreign affairs is a policy portfolio in which this division of labour seems to be already established: the Agency for Development Cooperation commissions about 90 per cent of all evaluations in that portfolio.

The fact that agencies commission far more evaluations than ministries means that evaluations are predominantly administered by organisations that are close to the policy under evaluation, not by a more remote, ultimate policy owner. This division of labour can bode well for the potential for realistic evaluation designs and for the utilisation of results from evaluations. However, it should be noted that, as policy implementation is often delegated to subnational government agencies and the local government sector, national executive agencies can often be relatively far removed from operation of the policy in question. Note also that, based on current data, we cannot distinguish whether commissioning an evaluation was the result of a unilateral agency decision or something that the ministry instructed the agency to do. A ministry can have political, as well as administrative, incentives to leave it to agencies to evaluate public policies.

The Evaluators

A total of 4,329 registered evaluators participated in one or more evaluations between 1994 and 2018 (see Table 1). The number of registered evaluators is higher than the number of actual evaluations because many evaluation projects/reports involved more than one evaluator. On average, each evaluator had been involved in eight to nine evaluations, with more than 100 evaluators involved in 10 or more. However, most of the 4,329 evaluators had only been involved in one or two evaluations.

Table 1. Top 10 evaluators by number of evaluations, 1994–2018

<i>Evaluator</i>	<i>Evaluations</i>
Office of the Auditor General	296
Norwegian Institute for Urban and Regional Research	235
Institute of Transport Economics	208
Pöyry and ECON	177
Nordic Institute for Studies in Innovation, Research and Education	141
Fafo Institute for Labour and Social Research	128
SINTEF	113
Vista Analyse	112
Norwegian Institute for Nature Research	106
Oxford Research	105
Total top 10	1621
Other evaluators	2708
Total	4329

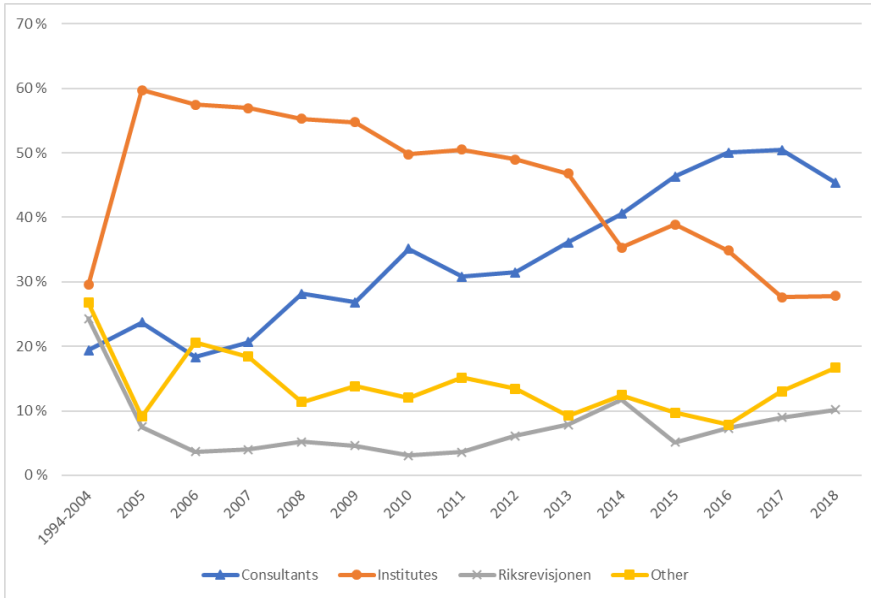
Source: Evalueringsportalen.no.

Table 1 displays the top 10 evaluators in Norway between 1994 and 2018, as measured by the number of evaluations. Taken together, these 10 performed 1,621 evaluations, or 37 per cent of all evaluations during the period. The top three are the Office of the Auditor General and two research institutes: the

Norwegian Institute for Urban and Regional Research and the Institute of Transport Economics. Pöyry, in fourth place, is the only non-Norwegian organisation in the top 10. This Finnish consultancy firm purchased the Norwegian consultancy ECON in 2008, and ECON’s evaluations are included in the figure for Pöyry. To sum up, the top 10 comprise three consultancy firms, six research institutes and the supreme audit institution.

In Figure 5, we grouped the evaluators into four categories: research institutes (including universities, colleges and Statistics Norway), private consultancies, the Office of the Auditor General, and “other” (including government organisations and ad-hoc committees).

Figure 5. Types of evaluators, percentages per year (N=4329)



Source: Evalueringsportalen.no

Evaluations recorded from 1994 through to 2004 are summarised for the sake of brevity, and we concentrated on the 2005–2018 period due to data comparability. A major shift occurred during this period, with consultants replacing research institutes as the government’s preferred evaluators. The proportion of evaluations conducted by research institutes was halved, from about 60 per cent in 2005 to about 30 per cent in 2018. The proportion done by consultants more than doubled, from around 20 per cent to around 50 per cent. For the Office of the Auditor General and the “other” category, the proportions were relatively stable during this period.

However, note that we measured evaluation volume as the number of evaluations. A different pattern could emerge if instead the volume was measured as the economic value of evaluations. We do not know whether systematic differences exist in size between evaluation projects conducted by research institutes and consultancy firms. However, it is likely that the Office of the Auditor General’s relative proportion would increase if the volume were measured using economic value.

Discussion

More research into and awareness of evaluation practices can be highly beneficial. An overview of evaluation practices might be the starting point for a summary of the knowledge inherent in thousands of evaluation reports: “A lot of social science is buried under rocks”, as Cook (1997) emphasised. The evaluation portal itself has improved the availability of evaluations in management. More knowledge of evaluation practices can also provide improved opportunities to identify high- and low-quality evaluation practices, clarifying what are good design choices for evaluations under different conditions and thus triggering a greater potential to produce evaluations that “speak truth to power” (Wildawsky 1987). High-quality evaluations provide a better knowledge base for policy decisions, and an evaluation’s quality can have an impact on its uptake in political decision-making (Weiss 1999). Employing the classification for evaluations proposed in this article can facilitate a systematic study of public evaluation practices, including a comparison of the evidentiary basis for policymaking across policy domains.

The analysis showed that the number of evaluations registered varied substantially over the 25-year period studied. The volume between 2010 and 2012 peaked at approximately 300 evaluations per year. These results only include external evaluations, so any evaluations performed in-house by, for example, government agencies themselves, are excluded from the analysis. One topic for future research is how the volume of internal evaluations has developed over time and how that volume might vary across policy areas. It would be particularly interesting to explore correlations with the use of external evaluations. In addition, we focus on the number of evaluations per year, not on the extent of government funding of evaluations. Any systematic change in the size of evaluations over time, for example in terms of their economic value, might produce a different impression of the more general trend in the volume of evaluations than our analysis has. It was not possible to obtain systematic data on the size of evaluation projects for this study. We note, however, that the Norwegian government commissioned an increasing number of major spending reviews in the 2010s, i.e. large-scale external evaluations, typically performed by consortia led by consultancy firms. The spending reviews cover a whole ministerial portfolio and produce several evaluation reports over a period of several years (Agency for Public and Financial Management 2015; see for example, Boston Consulting Group 2021). Further research could expand the dataset with more details about the size of evaluation projects in terms of funding, time frame and number of participants in consortia.

However, based on this article’s analysis of external evaluations registered in the Evaluation Portal, it appears that information regarding evaluation might have been less publicly available towards the end of the period studied than it was 10 years ago. This change in “evaluation intensity” might influence the information-based power possessed by different actors. Moreover, a change occurred in the commissioners’ relative composition. Executive agencies commissioned a larger proportion of the evaluations during the latter years of the period studied, during which the composition of evaluators also changed, with the consultants becoming more dominant than before. The agencies in the

Norwegian governance system are less powerful than the ministries, and the consultants may be less independent than the research institutes, which might affect the actors' capacity to set agendas. We discuss these three main findings below.

First, Norway's central government conducted a comprehensive evaluation of its own activities. Roughly 100 evaluation reports were delivered annually, though the number varied over time. Regular reports from the Office of Auditor General to Parliament represent a steady stream of evaluations. There are large differences in the scope of evaluation activities across ministries, as is evident in the number of reports. We suspect that this variation is partly due to traditions in the area and partly related to the predominant type of interventions in each policy area. Within defence and transport infrastructure, for instance, there are substantial irreversible investments that elicit major long-term effects and both seem to have rarely utilised evaluations. There are also areas dominated by time-limited interventions with a low level of resources that used evaluations more frequently as part of their activities. This is a somewhat paradoxical situation. The results presented in this article should be an eye-opener for the administration and political leadership. However, the results might have been intentional. The government, or at least the political elite, might not want its major political programmes to be evaluated and possibly questioned (Wildavsky 1987). Evaluations of time-limited activities that consume a relatively low level of resources might be less threatening to the system's status quo than the potential ramifications from evaluations of major, ongoing programmes and large-scale investments. The pattern that we have demonstrated in evaluation practices is, nevertheless, puzzling and presents a challenge for the research community to uncover causes and consequences of such differences in the scope and intensity of evaluations.

Second, the results show that in most policy areas the agency responsible for implementation orders the evaluation, rather than the higher-level authority (the ministry). When the subordinate agencies are the principal in evaluating their own policy interventions, there is less institutional distance between the actors responsible for the implementation of the measure than if the senior ministry had commissioned the evaluation, as most policies and policy instruments are implemented by government agencies or local governments. Potentially politically sensitive evaluations formally delivered to subordinate agencies, rather than to ministries, might be less threatening for the government or political elite than evaluations delivered formally and directly to a ministry. At the same time, the question of who is formally the *de facto* commissioner is a challenge for researchers to investigate. One interesting issue for further research is the degree to which institutional distance (separation of responsibility for commissioning evaluations and implementing policy measures) has implications for the evaluation's design, such as what type of evaluations are ordered.

Third, there was a notable shift on the part of the evaluators to having a larger proportion of evaluations conducted by consultants. External evaluations are mainly carried out by applied research institutes such as the Norwegian Institute for Urban and Regional Research or consultancy companies such as Pöyry. This presumably reflects the fact that evaluation projects are located in the grey area between contract research and advanced consultancy work. At the

same time, we observed an upward trend for consultancy companies and a corresponding downward trend for research institutes. One question that arises for further research is the degree to which this trend might be due to increasing price competition in the market for evaluation assignments or specialisation and professionalisation in the consultancy market. In absolute terms, the number of evaluations carried out by institutes has declined even more than the market share, while the number of evaluations commissioned from consultants has remained relatively stable, peaking in around 2010. Our results do not indicate why the consultants' market share has grown steadily. One partial explanation might be that consultancy firms expanded into this market during the evaluation "boom" around 2010, thereby intensifying competition during the subsequent decline. If price competition has become an issue, it is possible that research institutes, possibly due to more rigorous professional standards, have been neither able nor willing to engage in competition that would undermine research standards. Another possible explanation is that research institutes are less inclined to compete for smaller, short-term evaluation projects. Further research is needed to uncover the consequences of tendering processes for the selection of evaluators and the significance of price. As a result, competitive dynamics in the evaluation market may affect the validity of evaluation findings and the legitimacy of evaluations in the policy process.

The shift in the relative shares of evaluations between consultants and research institutes may pose consequences for the impact of evaluations. As consultancies operate according to a commercial logic, they might be less committed to professional standards (Skjøelsvik et al. 2016). This would imply that consultants are less independent than researchers in their framing of research problems, choice of methods and interpretation of findings. As a result, the trustworthiness of the evaluations' findings and the evaluators' legitimacy can be undermined. The government might consequently commission more evaluations from consultants because consultants might be more malleable to the government's needs, for example, delivering focused and timely evaluations. Accordingly, government agencies might assume that the consultants deliver more cost-effectively relative to the government's evaluation needs in the short-term. However, the differences between academic institutions and consulting firms may be less fundamental than suggested. For instance, major consulting firms are reportedly hiring an increasing number of PhDs (Venkatraman 2014), thereby presumably upgrading their methodological and study design skills as well as their connection to the national and international academic community. In addition, in some cases, research institutions are forming consortiums with consultancies in order to bid for evaluation contracts. If such hybridisation developments in the evaluation industry are taken onboard, further research could reach more nuanced conclusions about the type of evaluators that operate in the modern market for evaluations. Other questions worth pursuing in future research are the degree to which policy owners are inclined to prefer a certain type of evaluator, hoping for a favourable or convenient evaluation, and how such inclinations are conveyed to commissioners of evaluations.

If evaluators were to deliver evaluations that are politically sensitive, their independence and methods could potentially be questioned at a later stage by actors in the political arena. Such processes may give the government the option

of distancing itself from the evaluations. However, if the task of selecting evaluators and overseeing evaluation projects is entrusted to an actor independent of the policy owner, for example, the Research Council in the case of evaluations of welfare reforms or development aid, arm's length might be sufficient. On the one hand, evaluations resulting from such arrangements might be viewed as potentially less useful to the government in the short-term, as well as posing a higher risk of delivering politically sensitive evaluations than those commissioned directly by the government. On the other hand, evaluations commissioned at arm's length from the government can be time-consuming and, thus, when a politically sensitive evaluation is delivered, it might be too late to impact significantly on the programme or the political elite in question.

Another question might be whether this shift in relative proportion of evaluations between research institutes and consultants has implications for how evaluations are designed and conducted, for example, whether the consultants' preferred methods differ from those of researchers. A third question might be whether the trend of an increasing proportion of evaluations carried out by consultants is particular to Norway. Further research on evaluation practices should therefore track these developments. It would be interesting to see whether the trend of more consultant-based and fewer research-based evaluations continues or wanes; researchers should consequently examine the causes and possible consequences of such a development.

Conclusions

The analysis of more than 3,000 evaluations conducted for the Norwegian government during the 25-year period from 1994 to 2018 revealed three patterns. First, about 100 evaluations were registered in the government evaluation database annually. It is likely that there was some under-reporting in general, and in particular for the later years. The analysis of the data from the Evaluation Portal warrants a cautious conclusion that the volume of evaluations in fact peaked around 2010–2012 and has since declined. Triangulation with supplementary data sources also indicates that the total volume has decreased since the early 2010s. In any case, given that there has been a decline in the evaluation reports available via public platforms and libraries, there is less information available in the public sphere about the success and failure of public policies in the late-2010s than there was earlier.

Second, ministries have commissioned relatively fewer evaluations in the last decade than in the years before, and executive agencies have commissioned relatively more. Third, consultants' share of evaluations has risen, while that of research institutes has fallen. These tendencies might be explained by an increased demand for more timely and focused evaluations than may have been conducted in the "heydays" of the earlier waves of evaluation. However, these tendencies may also be explained by shifts in power relations between major actors in the evaluation industry and polity. These latter two conclusions are vulnerable to limitations in the data material. However, we have no indication from supplementary sources that ministries have been more likely than government agencies to under-report evaluations to the database; neither do we

have any indication that evaluation reports from research institutes are less diligently reported in the database than those from consultants.

The study has some limitations, not only those that concern the data source, but also in the choice of research design which focuses on the number rather than the size of the evaluations. We encourage future research that takes a systematic look at developments in the economic size of the evaluations and, relatedly, in the complexity of evaluation projects. Important trends can occur outside the view of studies focused on the number of evaluations, for example, that government organisations develop a preference for commissioning fewer and larger evaluations, similar to developments observed in the public funding of research (Bloch and Sørensen 2015). While we were able to compile longitudinal data, time series even longer than ours are needed to discern long-term developments associated with the ebbs and flows of politics. Further research based on richer data and longer time series may accordingly be needed to reach firm conclusions about the development in the volume of evaluations in recent decades. Another limitation is that, although the paper covers methods used in the evaluations, the analysis of those methods says little about the quality of evaluations. We encourage future research that looks more closely at the quality of evaluation projects and reports, for example, by comparing them with benchmarks as suggested by Vedung (2005).

Evaluations may be useful for many stakeholders, but producing the evaluations is costly. The design of the evaluations, specifically how the commissioners choose the type and timing of the evaluations and how the evaluators choose methods and modes of communication for the results, could have a major impact on the costs and benefits of the evaluations. The systematisation and analysis of evaluations, such as in this paper, can provide the basis for a more efficient evaluation practice in the public sector. For instance, we show that there are notable differences in the scope of evaluation activities across policy areas, seemingly unrelated to the scope and significance of policy areas. Numerous or frequent evaluations, as seen in some policy areas, do not necessarily provide better information or knowledge bases. Small-scale evaluations with a short time horizon provide a relatively weak basis for conclusions about how well a measure or a scheme is working. A smaller number of evaluations, but larger-scale and longer-term, usually provide safer conclusions and may thus be preferable to many stakeholders though not produced at the right time for decision-making relevant to the programme in question.

The design of the evaluations is evidently subject to cost-benefit considerations, but may also be subject to vested interests and the power of major stakeholders. In policy areas characterised by relatively small and time-limited policy interventions, the evaluations will presumably be similarly small scale in terms of costs and time-span, and this might be an explanation for the frequent evaluations in aid and regional business development. One practical implication could be to design interventions with a longer horizon and to commit more resources to allow more thorough evaluations. There is also a tendency for small and short-term evaluations to focus on, or draw the safest conclusions about, immediate effects and practical aspects of implementation. Such information can be very useful for the agencies, but less useful for policy makers

or the public. This area would benefit from further research into the power of evaluations and the consequences of types of evaluations and methods used across policy fields.

Acknowledgements

We acknowledge the constructive comments we received when an early version of this paper was presented for the Network for Organizational Research in Norway (NEON) annual conference, Oslo, 18–19 November 2020. We also acknowledge the constructive comments from the editors and the anonymous reviewers.

References

- Agency for Public and Financial Management (2015) *Bruk av sektoranalyser og evalueringar. Ei kartlegging*. Rapport 3/2015. The Agency for Public and Financial Management, Oslo.
- Andersen, V. N., & C. S. Pedersen (2012) Evaluering af (og i) i den danske offentlige skole, *Nordiske Organisasjonsstudier*, 14 (1): 3–27.
- Askim, J., E. Døving & Å. Johnsen (2013) 'Evalueringsspraksis i den norske statsforvaltningen 2005–2011' in A. Halvorsen, E. L. Madsen & N. Jentoft (eds.), *Evaluering: Tradisjoner, praksis, mangfold*, pp. 72–85, Fagbokforlaget, Bergen.
- Askim, J., K. H. Bjurstrøm & J. Kjærvik (2019) Quasi-contractual ministerial steering of state agencies: Its intensity, modes, and how agency characteristics matter, *International Public Management Journal*, 22 (3): 470–498.
- Baklien, B. (1993) Evalueringsspraksis i Norge, *Tidsskrift for samfunnsforskning*, 34: 261–274.
- Balthasar, A. (2009) Institutional design and utilization of evaluation: A contribution to a theory of evaluation influence based on Swiss experience, *Evaluation Review*, 33 (3): 226–256.
- Bloch, C. & M. P. Sørensen (2015) The size of research funding: Trends and implications, *Science and Public Policy*, 42 (1): 30–43.
- Boston Consulting Group (2021) *Områdegjennomgang av styringen av politi- og lensmannsetaten*, Rapport for Finansdepartementet og Justis- og beredskapsdepartementet, Boston Consulting Group, Oslo.
- Bovens, M., P. T. Hart & S. Kuipers (2006) 'The politics of policy evaluation' in M. Moran, M. Rein and R.E. Goodin (eds.), *The Oxford Handbook of Public Policy*, Oxford University Press, Oxford.
- Profoss, K. E. (1997) 'Departementenes evalueringsspraksis' in Statskonsult (ed.), *Evalueringsspraksis i departementa: Rapport frå Statskonsult sin konferanse om evaluering, april 1997*, kapittel 5, Rapport 1997:13, Statskonsult, Oslo.
- Christie, C. A. & Alkin, M. C. (2013) 'An evaluation tree' in M. C. Alkin (ed.), *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences*, 2nd ed., pp. 11–57, Sage, Thousand Oaks, California.

- Christie, C. A. & Fleischer, D. N. (2010) Insight into evaluation practice: A content analysis of designs and methods used in evaluation studies published in North American evaluation-focused journals, *American Journal of Evaluation*, 31 (3): 326–346. doi.org/10.1177/1098214010369170
- Cook, T. D. (1997) ‘Lessons learned in evaluation over the past 25 years’ in Eleanor Chelimsky & William R. Shadish (eds.), *Evaluation for the 21st Century: A Handbook*, Sage, Thousand Oaks.
- Coryn, C. L., Wilson, L. N., Westine, C. D., Hobson, K. A., Ozeki, S., Fiekowsky, E. L., Greenman, G.D. & Schröter, D. C. (2017) A decade of research on evaluation: A systematic review of research on evaluation published between 2005 and 2014, *American Journal of Evaluation*, 38 (3): 329–347.
- Dahler-Larsen, P. (2000) Surviving the routinization of evaluation, *Administration & Society*, 32 (1): 70–92.
- Eriksen, B. (1990) ‘Institutional aspects of evaluation in Norway’ in Ray C. Rist (ed.), *Program Evaluation and the Management of Government. Patterns and Prospects across Eight Nations*, Transaction Publishers, London.
- Furubo, J. E., Rist R. C. & Sandahl R. (ed.) (2002) *International Atlas of Evaluation*, Transaction Publishers, New Brunswick, N.J.
- Furubo, J. E. (2007) ‘Policy Analysis and Evaluation in Sweden: Discovering the Limits of the Rationalistic Paradigm’ in F. Fischer, G.J. Miller & M. S. Sidney (eds.), *Handbook of Public Policy Analysis*, pp. 571–586, CRC press, Boca Raton/London/New York.
- General Accounting Office (1990) *Prospective evaluation methods. The prospective evaluation synthesis*, GAO/PEMD-10.1.10. General Accounting Office, Washington, D.C.
- Hansen, H. F. (2003) *Evaluering i staten: kontrol, læring eller forandring?*, Samfundslitteratur, Frederiksberg.
- Henry, G. T. & Mark, M. M. (2003) ‘Toward an agenda for research on evaluation’ in C. A. Christie (ed.), *The practice–theory relationship in evaluation*, pp. 69–80, Jossey-Bass, San Francisco.
- Howell, E. M. & Yemane, A. (2006) An assessment of evaluation designs, *American Journal of Evaluation*, 27 (2): 219–236.
- Jacobs, S., Speer, S. & Furubo, J. E. (2015) The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later, *Evaluation*, 21 (1): 6–31.
- Kallemeyn L.M., Hall J., Friche N. & McReynolds C. (2015) Cross-continental reflections on evaluation practice: Methods, use, and valuing, *American Journal of Evaluation*, 36 (3): 339–357.
- Leeuw F. L. & Furubo J. E. (2008) Evaluation systems: what are they and why study them?, *Evaluation*, 14 (2): 157–169.
- Leeuw, F. L. (2009a) Evaluation: a booming business but is it adding value?, *Evaluation Journal of Australasia*, 9 (1): 3–9.
- Leeuw, F. L. (2009b) Evaluation policy in the Netherlands, *New Directions for Evaluation*, Vol. 123: 87–102.
- Leeuw, F.L. & P.J. Rozendal (1994) ‘Policy evaluation and the Netherlands’s government: Scope, utilization and organizational learning’ in F. L. Leeuw,

- R. C. Rist & R. C. Sonnichsen (eds.), *Can Governments Learn?*, New Brunswick, NJ: Transaction Publishers.
- Lemire, S., Fierro, L. A., Kinarsky, A. R., Fujita-Conrads, E., & Christie, C. A. (2018a) The U.S. federal evaluation market, *New Directions for Evaluation*, 160: 63–80.
- Lemire, S., Nielsen, S. B., & Christie, C. A. (2018b) Toward understanding the evaluation market and its industry—Advancing a research agenda, *New Directions for Evaluation* 160: 145–163.
- Lemire S., Peck, L. R. & Porowski, A. (2020) The growth of the evaluation tree in the policy analysis forest: Recent developments in evaluation, *Policy Studies Journal*, 48 (S1): S47–S70.
- Løwendahl, B. (2016) ‘Strategi og ledelse i profesjonsdominerte virksomheter’ in E. Døving, B. Elstad and Aa. Storvik, *Profesjon og ledelse*, pp. 99–119, Fagbokforlaget, Bergen.
- Nielsen, S. B., Lemire, S., & Christie, C. A. (2018) The evaluation marketplace and its industry, *New Directions for Evaluation* 160: 13–28.
- Norwegian Agency for Public and Financial Management (2020a) Om evalueringsportalen. Web page accessed 23.12.2020. <https://evalueringsportalen.no/om-evalueringsportalen>
- Norwegian Agency for Public and Financial Management (2020b) Evalueringsportalen. Web page accessed 23.12.2020. <http://evaluering.nb.no/>
- Nutley, S. M., I. Walter & H. T.O. Davies (2007) *Using Evidence: How Research Can Inform Public Services*, Policy Press, Bristol.
- Pawson, R. (2006) *Evidence-Based Policy: A Realist Perspective*, Sage, London.
- Pfeffer, J. & R. I. Sutton (2006) *Hard Facts, Dangerous Half-Truths, and Total Nonsense*, Harvard Business School Press, Boston, Mass.
- Power, M. (1997) *The Audit Society: Rituals of Verification*, Oxford University Press, Oxford.
- Research Council of Norway (2018) ‘Godkjente forskningsinstitusjoner’. <https://www.forskningsradet.no/sok-om-finansiering/hvem-kan-soke-om-finansiering/forskningsorganisasjoner/godkjente-forskningsorganisasjoner/>
- Research Council of Norway (2019) ‘Statlig grunnfinansiering av forskningsinstitutter og forskningskonsern’. <https://www.forskningsradet.no/om-forskningsradet/oppgaver-organisering/instituttsektoren/statlig-basisfinansiering-av-forskningsinstituttene/>
- Sanderson, I. (2002) Evaluation, policy learning and evidence-based policy making, *Public Administration*, 80 (1): 1–22.
- Scriven, M. (1991) *Evaluation Thesaurus*, 4th ed. Sage, Newbury Park, California.
- Skjølsvik, T., K.J. Breunig and F. Strønen (2016) ‘Hjelp, ledelsen kommer!’ in E. Døving, B. Elstad and Aa. Storvik, *Profesjon og ledelse*, pp. 77–97, Fagbokforlaget, Bergen.
- Vedung, E. (2005) *Public Policy and Program Evaluation*, Transaction Publishers, London.

- Vedung, E. (2006) 'Evaluation research' in B.G. Peters & J. Pierre (eds.), *Handbook of Public Policy*, Sage, London.
- Vedung, E. (2010) Four waves of evaluation diffusion, *Evaluation*, 16 (3): 263–277.
- Venkatraman, V. (2014) The science careers guide to consulting careers for Ph.D. scientists. *Science Magazine*, September 3, 2014, accessed April 13, 2014. <https://www.sciencemag.org/careers/2014/09/science-careers-guide-consulting-careers-phd-scientists>
- Vining, A. R. & D. L. Weimer (2006) 'Efficiency and cost-benefit analysis' in B.G. Peters & J. Pierre (eds.), *Handbook of Public Policy*, Sage, London.
- Von Nordenflycht, A. (2010). What is a professional service firm? Toward a theory and taxonomy of knowledge-intensive firms, *Academy of Management Review*, 35 (1): 155–174.
- Weiss, C. H. (1987). Evaluating social programs: What have we learned? *Society*, 25 (1): 40–45)
- Weiss, C. H. (1999) The Interface between Evaluation and Public Policy, *Evaluation* 5 (4): 468–86.
- Wildavsky, A. (1987) *Speaking Truth to Power: The Art and Craft of Policy Analysis*, 2nd edition, Transaction Publishers, New Brunswick.
- Wærness, M. & R. Øvrelid (2003) *Departementenes bruk av evalueringer*. Notat 2003:1, Statskonsult, Oslo.
- Øvrelid, R. (2000) Evaluering i Norge, *Nordisk Administrativt Tidsskrift*, 81 (2): 131–147.