

Understanding the use and usability of research evaluation studies^{1,2}

Katrin Milzow^{1,*}, Anke Reinhardt², Sten Söderberg³ and Klaus Zinöcker⁴

¹Division for Strategy Support, Swiss National Science Foundation (SNSF), Bern 3001, Switzerland, ²Deutsche Forschungsgemeinschaft (DFG), Information Management Group, Bonn 53170, Germany, ³Department of Research Policy, Swedish Research Council (VR), Box 1035, Stockholm 10138, Sweden, and ⁴Department for Policy, Evaluation, Analysis, Science Fund (FWF), Sensengasse 1, Vienna 1090, Austria

*Corresponding author. Email: katrin.milzow@snf.ch

Sadly, one of the authors, Klaus Zinöcker, passed away suddenly on 3 March 2018. He was involved in all stages of the research and production of this article and gave his assent to the submission of the article to *Research Evaluation*.

In memory of Klaus Zinöcker

Abstract

Many research funding and research performing organizations evaluate their operations or are subject to evaluations. A Working Group within Science Europe undertook to illustrate how *ex post* evaluation of research is actually used. Based on 11 case studies from seven organizations and countries, we explore the use of evaluation studies commissioned or conducted by member organizations. We draw on a theoretical framework that categorizes types of evaluation use, including legitimation, information, learning, steering, mediating, tactical, and ritual use. To explore the factors influencing the use of evaluations, we draw on a 'logic model of evaluation use' developed by Mark and Henry (2004). The study shows that many uses of evaluation studies can (co-)exist. In particular, direct influence on decision-making (steering use) is not the only type of use that practitioners see in their daily work. Equally important types of use include contribution to the knowledge base and cognitive resources within organizations and beyond. Our study highlights that a methodologically rigorous and high-quality evaluation study is a necessary, but not sufficient, condition to use evaluation results. It draws attention to the fact that the context and other factors external to the evaluation itself—but often within the realm of the commissioning organization—determine the actual use of the evaluation. The findings suggest applying logic models not only to evaluation objects but also to the evaluation exercise itself, so as to consider the underlying mechanisms and chains of influence through which evaluation may have its effects.

Key words: logic model; *ex post* evaluation; research funding; stakeholder; involvement; evaluation influence.

1. Introduction

1.1 Rationale and objectives

Research funding and research performing organizations actively engage in *ex post* evaluation activities. Their underlying rationale is—amongst others—to provide an evidence base for strategy development, establish transparency and accountability in their activities, support internal learning processes, and communicate

with stakeholders.³ The field of evaluation of research has grown rapidly and while many organizations have developed evaluation standards and processes to guarantee professionalism and quality⁴, attention turns to the question how evaluation is actually used.⁵

A Science Europe⁶ Working Group on Evaluation—consisting of evaluation specialists who manage evaluation projects and work at the interface of evaluators and decision-makers—undertook to

document and analyse the uses of evaluation studies commissioned or conducted by their respective research or research funding organizations. The study brings in a practitioners' perspective on aspects of evaluation use that emerge in the specific context of research evaluation in Europe. It is motivated by the manifold, sometimes conflicting expectations concerning the use of evaluation of research that have been observed by members of the Science Europe Working Group in their daily practice. We rely on existing conceptual models⁷ of use and test them with examples from a wide range of organizations, countries, and circumstances. Through case studies, we attempt to illustrate which types of use (co)existed in a variety of research evaluation settings and to analyse which factors facilitated or, respectively, were detrimental to the use and usability of evaluation studies.

Based on our practical experience at the interface between evaluators, programme managers and other stakeholders, we attempt to:

- better understand the role *ex post* evaluation plays in the research (funding) system;
- identify different patterns of evaluation use and conditions under which evaluations can fulfil their respective intentions—from contributing to the knowledge base, to policy learning, legitimization, etc.; and
- draw lessons learned and explore whether recommendations can be provided on this basis to ensure evaluations are targeted and fruitfully employed.

In short, the objective is to understand the role of evaluation and use this knowledge to make evaluation worthwhile within the field of research evaluation—and hopefully also in a broader context. We would not wish to suggest that increasing the use of evaluation is a good in itself. Rather the potential use of evaluation outputs (and of evaluation) should be assessed critically within organizations engaging in research evaluation activities.

1.2 Relation to other work

We use extant theory as a guide in our analysis and understanding of the problems of use and usability that emerge in the specific context of research evaluation in Europe. We take an approach similar to Streicher, who recently examined the effects of evaluations in the Austrian research, technology, and innovation system, but expand the empirical basis to include several countries and types of organizations.⁸ As a starting point, we build on Evert Vedung's definition of evaluation. According to Vedung, '[evaluation is the] careful retrospective assessment of the merit, worth, and value of administration, output, and outcome of government interventions, which is intended to play a role in future, practical action situations'.⁹ The rationale for choosing this definition of evaluation is that we focus on *ex post* or mid-term evaluation—in contrast for example to *ex ante* evaluation, which is common practice in research funding agencies.

Evaluation use is 'arguably the most studied area of evaluation', and at the same time studies show that evaluations frequently remain unused.¹⁰ In a seminal work published in 1979, Carol Weiss asked, 'what "using research" actually means'.¹¹ She distinguished, amongst others, between the 'problem-solving model', where the results of a study, or the findings and recommendations of an evaluation, are applied to a 'pending decision' or to 'help solve a policy problem', and the 'enlightenment model', where 'concepts and

theoretical perspectives (...) permeate the policy-making process'.¹² Weiss suspected that the latter 'is the way in which social science research most frequently enters the policy arena', whereas the first model 'probably describes a relatively small number of cases'. The distinction is relevant to our case studies and can be framed in terms of 'instrumental' (or steering) and 'conceptual' use, as done by Weiss in a later paper. Other authors, such as Christina Boswell and Lorna Schrefler, distinguish instrumental use of evaluation from symbolic use, such as legitimizing functions, as well as strategic use, where evaluation contributes to the legitimacy of an organization or the substantiation of particular policy choices.¹³

Our case study design draws on the types of use proposed specifically for research evaluation by the Austrian Platform Research and Technology Policy Evaluation (fteval) in its 'Evaluation Standards in Research and Technology Policy',¹⁴ which reflects the above distinctions.

These types of uses include:

- **Legitimation:** e.g. justifying the use of public funds
 - **Information**¹⁵: providing decision-makers or the public with information on how public funds are being used and to what effect
 - **Learning:** for funders, decision-makers in the field of science, innovation or technology policy, intermediaries, universities, research institutions, enterprises, consultants, and scientists
 - **Steering:** for establishing policy objectives and planning of measures
 - **Mediating:** between the competing interests of different players
- We draw on Evert Vedung¹⁶ for two additional types of use:

- **Tactical**¹⁷: to gain time or avoid responsibility
- **Ritual**¹⁸: systematic evaluation at certain intervals/on certain occasions.

On a general level, the empirical findings confirm that these categories are practical for the present purpose: the instances of use identified can be attributed to one of the types, which are reasonably clearly delimited and for the most part relevant to the case studies.

Furthermore, the analysis is informed by the many studies attempting to highlight factors facilitating the use of evaluation, including the soundness of methods, stakeholder involvement in the evaluation process, and other human and contextual factors.¹⁹ Many authors highlight the significance of context²⁰ and that evaluators become strategic actors, endogenous to the evaluation process.²¹ Our classification of factors facilitating/hindering evaluation use draws on the 'logic model of evaluation use' developed by Mark and Henry. They describe the logic model as 'a framework designed to capture key mechanisms through which evaluation may have its effects', in analogy to 'the logic models or program theories that many evaluators apply to the programs they are evaluating'.²²

1.3 Methodology

Only retrospective, completed, and published evaluation studies are included in the current analysis (interim or *ex post*, i.e. following Vedung's definition of evaluation). As an additional delimiting element for the selection of case studies, we include only studies that were performed independently, which is frequently considered an important quality criterion.²³ Independence in our context means that the studies either were commissioned to external contractors or included external committee members. The case studies cover evaluation objects at the level of the institution, research field, or instrument/programme but not individual grants, persons, or projects.²⁴

These evaluations have taken place at the organizations represented in the Science Europe Working Group. Thus, biases in their selection cannot be excluded.

To enable the exploration of potential factors influencing evaluation use, we selected case studies to cover various degrees of use, from extensive to less extensive use. The case studies were thus selected based on the dependent variable, i.e. the use of the evaluation. While this approach may not necessarily be representative, it can help explore various factors that may affect the use of evaluation.²⁵

The case studies were carried out and written by members of the Science Europe Working Group in accordance with an agreed case study data collection template, which had been tested and refined in a pilot study. Eight persons from seven organizations in different countries documented the 11 case studies that constitute our empirical material.

In addition to different kinds of publicly available written information, the case studies are informed by interviews with stakeholders contributing different perspectives to the evaluation: representatives of the commissioning instances, programme managers for the research programmes evaluated, and/or evaluators/evaluation managers. In all cases, at least two inputs (interview and/or written material) from different viewpoints were used. The number of persons interviewed for each case study varied and was largely dependent on the degree of involvement or on logistic reasons such as availability. A total of 25 persons were interviewed, varying between one person (in three case studies) and up to five persons (one case study). The interviewees address issues on evaluation use based on pre-defined open questions, according to the above-mentioned template for data collection (see Appendix 2). The questions cover the process through which evaluations exerted influence, the stakeholders effected by the evaluation, and the ways through which they were effected.

The case studies were then analysed and interpreted independently by at least two persons from the Science Europe Working Group, following a specific analysis template (see Appendix 3). In this stage, evidence for different types of evaluation use and different types of factors influencing use was systematically identified and structured.

It should be noted that the methods used for data collection—use of publicly available information plus interviews with key actors involved in and/or affected by the evaluation exercise—imply that we cannot exclude that there may be unobserved effects from the evaluations. These effects could include, for example, third parties taking up evaluation reports or effects occurring after the date of the interview. Such effects can be extremely difficult to detect and could, of course, alter the overall cost–benefit associated with an evaluation.

There is a substantial degree of heterogeneity between three types of case studies displayed in Table 1. Case Study 1 considers evaluations of research units by the French Research Evaluation agency Haut Conseil de l'Évaluation de la Recherche et de l'Enseignement Supérieur (HCERES). Case Studies 2–5 deal with the evaluation of research fields, with national funding agencies as evaluators or coordinators of the evaluation. In Case Studies 6–11, the evaluation object was the funder or a funding programme, evaluated by an external entity. Whereas the evaluations of research fields and institutions tend to be recurrent, evaluations of instruments or programmes tend to be situation-specific *ad hoc*. The inclusion of different evaluation objects allowed us to cover a wider range of potential uses and to draw on various aspects in the field of research evaluation.

The method used and the decisions that guided the selection of cases restrict the possible outcome of the analysis. Even though we illustrate the empirical results quantitatively, we do not aim for quantifiable conclusions. Our approach also does not allow for describing the phenomenon under consideration exhaustively or claiming causal relationships. Instead, we aim to point out different constellations of evaluation studies and evaluation uses that occur in practice.

2. Findings

2.1 How were evaluation studies used?

The analysis highlights that many uses of evaluation studies can and do (co-)exist. All evaluations had multiple uses. As apparent from Fig. 1, 'steering' and 'information' use were most widespread, followed closely by 'learning use'. 'Legitimizing' and 'ritual use' were noted in just over half of the case studies. There were no examples of 'tactical use' and little evidence of 'mediating use'.

'Legitimizing use', to justify the use of funds, goes along with other uses in all case studies, as evaluations are expected to fulfil both summative and formative functions. Whereas 'legitimizing use' is more common for evaluations of instruments and programmes, 'ritual use' (the systematic evaluation at certain intervals or on certain occasions) is noted mainly for evaluations of research institutions and fields. 'Ritual use' is associated with 'learning' more than 'legitimising' functions. The case studies showing 'ritual' use also consistently indicate that 'steering' use was very relevant. This could be because recurrent evaluations are more embedded institutionally (see Chapter 2.2). Evaluations of research fields and institutions frequently aim explicitly for learning effects for individual research institutes or units and accordingly present detailed results and recommendations for different units. For example, the Evaluation of Mathematics in Norway included an assessment of the value gained from recommendations from previous evaluation in strategic planning by the institutions (Case Study 4).

Total 10 out of 11 case studies show so called 'steering use', where evaluations supported decision-making or led to policy change. Evaluations have led to changes to funding instruments, as in the case of the evaluations of Mathematics in Norway (Case Study 4) and the Swiss National Science Foundation (SNSF) Fellowship Scheme (Case Study 11), or facilitated the fusion of several instruments as with the Evaluation of Education Programmes in Norway (Case Study 9). Additional funds for new initiatives were made available following the evaluations of Mathematics in both Sweden and Norway (Case Studies 3 and 4). Evaluations of institutions or research fields have supported decisions about the structure of research units or institutes in the case of the Evaluation of Plant and Soil Sciences in Estonia (Case Study 5) or the evaluations by the French Research Evaluation agency HCERES (Case Study 1).

The case of Mechanical Engineering in Sweden (Case Study 2) illustrates that 'steering use' may also manifest itself in non-action, if the evaluation prevents a planned or discussed 'negative action', for example an economic cutback of a certain research program or research field.

Alongside 'steering use', 'information' and 'learning use' were most frequently noted in the case studies. To some extent, information and learning use are prerequisites to steering use; however, they can also exist independent of each other. The evaluation of the Deutsche Forschungsgemeinschaft (DFG)'s Emmy Noether

Table 1. Overview of case studies

Evaluation object	Case study	Evaluator	Date of publication
Institution	Evaluation of INRA-Research Units by HCERES	HCERES	Several evaluations
Research field	Swedish Research in Mechanical Engineering	International expert committee	2013
	Swedish Research in Mathematics	International expert committee	2010
Instrument/ Programme	Research in Mathematics at Norwegian Universities	International expert committee	2012
	Evaluation of Plant and Soil Sciences in Estonia	Foundation Archimedes, International expert panel	2012
	DFG Collaborative Research Centres	Joanneum Research	2012
	DFG Emmy Noether Career Programme	Institut für Qualitätssicherung (IfQ)	2008
	Swedish Research in Biodiversity	Two expert committees for Scientific Quality and for Relevance	2010
	Education Programmes in Norway	Panel of experts in educational science from the Nordic countries	2013
	National Prevention Research Initiative MRC-UK	Independent Scientific Review Group	2015
	SNSF Fellowship Schemes	Observatoire Science, Politique et Société, University of Lausanne	2010

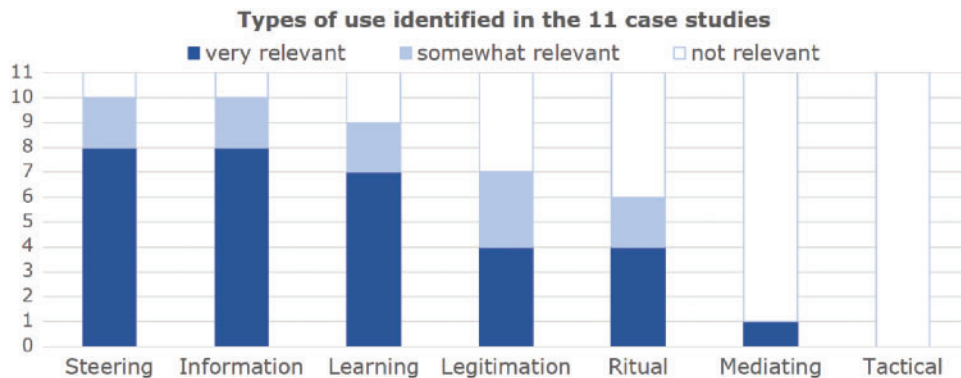


Figure 1. Types of use identifies in the 11 case studies.

Programme (Case study 7) did not directly lead to policy change but generated substantial interest and impact in the research policy community. It was downloaded already more than 1,500 times shortly after the publication. This outcome highlights that in the absence of ‘steering use’, evaluations can and do contribute to the knowledge base and cognitive resources within organizations and beyond.

Evaluation studies also have indirect effects beyond the institution, the research field, or the programme evaluated. Methodological cross-fertilization with other evaluation studies is noted on several occasions. Lessons learned from the Evaluation of Swedish Research in Mathematics were applied to the Evaluation of Mechanical Engineering (Case Studies 3 and 2, respectively). The evaluation of Swedish Biodiversity Research was considered innovative for involving two expert committees for Scientific Quality and Relevance and subsequently influenced several other evaluation studies (Case Study 8). In the case of the Evaluation of Mathematics in Norway, too, learning on how to do subject-specific evaluation was noted as a form of use (Case Study 4).

The prevalence of indirect effects suggests that in several cases, it is relevant to talk about the influence of evaluation studies rather than the use—the latter term suggests a more direct or instrumental relationship between an evaluation study and its effects. Since our

cases studies are based on the notion of ‘use’, we continue to use this term in this article.

2.2 What supports or hinders the use of evaluation studies?

When looking at circumstances that may facilitate evaluation use or be detrimental to it, the context of the evaluation seems to play a larger role than the actual evaluation activities or their immediate inputs and outputs. This is consistent with the current academic debate.²⁶

Using the logic model framework developed by Mark and Henry,²⁷ we identified context as relevant to the use of evaluation results in all 11 case studies, as shown by Fig. 2. Context is also the factor most frequently identified across all case studies, as shown by Fig. 3, which demonstrates the types of factors highlighted by the case studies and their frequency, indicated by the length of the arrows. The actual evaluation activities were mentioned in only 5 of 11 case studies.

2.2.1 Contextual factors

Political circumstances at the level of the organization or the research landscape (i.e. micro- and macro politics) play an important

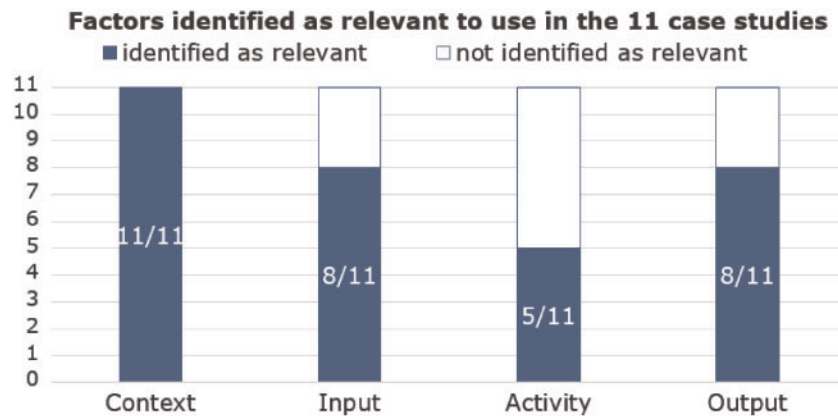


Figure 2. Factors identified as relevant to use in the 11 case studies.

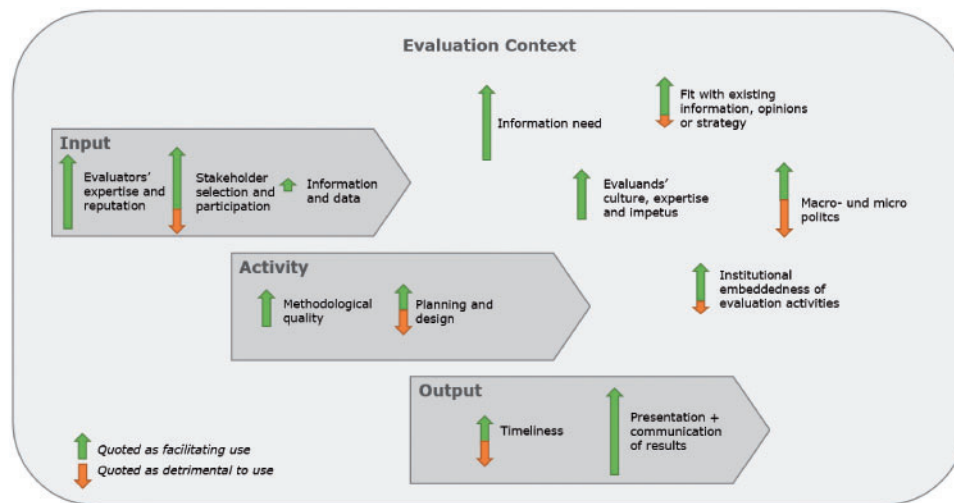


Figure 3. Factors influencing use identified in the case studies. Representation inspired by Mark and Henry (2004).

role in our case studies. The evaluation of Educational Programmes in Norway followed a policy for innovation in the public sector, which was issued by the Research Council of Norway during the previous year (Case Study 9). In the case of the evaluation of Mathematics in Sweden, the subject field of mathematics was already widely discussed in the context of the results of international comparative studies of student performance in schools (TIMSS and PISA), which meant that the policymakers, the research community, and the general public in Sweden were aware of issues affecting competence, education, and (to some extent) research in Mathematics—this issue was discussed as a ‘Mathematics Crisis’ in Sweden at the time. Thus, the political context provided a fertile environment for initiatives as a consequence of the evaluation (Case Study 3). Evaluations mandated by ministries (e.g. the evaluation of educational programmes in Norway, Case Study 9) as well as instances where a clear political will for change was communicated (see Case Study 5, Plant and Soil Sciences in Estonia) also increase the likelihood that an evaluation leads to change.

Political circumstances may also work to limit the use of an evaluation. Budget cuts prevented the implementation of some recommendations in the case of the evaluation of DFG Collaborative

Research Centres (Case Study 6). In the case of the follow-up study to the evaluation of the UK-National Prevention initiative, too, the implementation of recommendations is perceived to depend above all on budgetary decisions (Case Study 10). In the case of the evaluation of Biodiversity in Sweden, political groups calling for these specific programmes for biodiversity research had left government, and the respective funding had already dried up before the evaluation could be completed (Case Study 8). Accordingly, political changes had reduced the relevance of the resulting recommendations.

This shows that it is important to consider context before investing resources into an evaluation. Even in the absence of obviously impeding factors, political will to implement findings is required. This may not always be possible, since circumstances may change quickly and unexpectedly.

One of the factors promoting the use of evaluation results quoted most frequently in the case studies is that the evaluation addressed information needs or knowledge gaps; for example, when there were concerns over the quality of the evaluation object or in relation to difficult policy decisions (e.g. budget decisions and restructuring). This finding is evident in the evaluation of Plant and Soil Sciences in

Estonia, where commissioning parties saw the need for restructuring the research field and accordingly had a strong interest in designing an evaluation, which could provide practical and useful results (Case Study 5). Similar configurations can be observed in the evaluation of Mechanical Engineering in Sweden (Case Study 2), the Evaluation of the UK's National Prevention Research Initiative (Case Study 10), and the Evaluation of Educational programmes in Norway (Case Study 4). At the same time, the lack of a perceived need for urgent reform may result in incremental recommendations or non-implementation of recommendations (Case Studies 11 and 7).

A strong interest in the results of an evaluation may go along with specific expectations and may thus be a threat to the independence of the evaluation—or lead to disappointment with the results.

Two case studies highlight the fit with existing information, opinions, or strategy as a factor determining evaluation use. The Evaluation of Biodiversity in Sweden added to several pieces of information pointing in a similar direction for one of the commissioning parties, where some steering use was noted. The recommendations fit poorly with the general strategy of the second commissioning party, where no steering use was noted. The informants for the case study indicated that the mismatch between the recommendations and strategic orientation might have played a role in this case (Case Study 8). In the case of the SNSF Fellowship programme, recommendations were in line with existing ideas, which facilitated their implementation (Case Study 11).

Another example of contextual factors, which played a role, and which might counterbalance particular interests, was the institutional embeddedness of evaluation activities. In the case of Norway, most evaluations are followed up with an allocation of funds for strategic projects. 'Follow-up committees' with representatives from research institutions consolidate follow-up plans, and the respective evaluation accordingly included an assessment of the degree to which previous evaluations had been used by institutions for their strategic planning (Case Study 4). In France on the other hand, there are some concerns that insufficient coordination of evaluations at many levels (individual researchers, projects, research units, and institutions) may reduce their impact (Case Study 1).

2.2.2 Evaluation inputs, activities, and outputs

Among factors linked more directly to the evaluation itself, the evaluation inputs were indicated to have facilitated evaluation use in 8 out of the 11 case studies. The evaluator's expertise and reputation was mentioned in six case studies. It is possible that this was relevant in other cases too but considered an obviously necessary condition for evaluation use and therefore not mentioned explicitly. In any case, it seems that today,²⁸ and in relation to our case studies and the organizations considered therein, this is not a major issue of concern.

Stakeholder selection and participation, which links back to the context of the evaluation, was considered instrumental to evaluation use in a majority of cases. With all evaluations of research fields and research institutes, the researchers were involved in some way. They for example had the opportunity to suggest members for the evaluation panels (Case Studies 3 and 4), were asked to self-report (Case Studies 2 and 4), or included in decisions about the evaluation design (Case Study 5). Thus involved, institutions in Norway viewed the evaluation of research in Mathematics as a possibility to get more funds and possibly new programmes for their field. However, not all types of institutions chose to participate in this evaluation

(Case Study 4). In the case of the evaluation of Swedish Research in Mathematics, two funding bodies initiated the evaluation together, which was considered to add to the substance of the results and increase the probability that they will be used (Case Study 3).

Insufficient stakeholder involvement was considered a possible impediment to use by stakeholders at INRA in certain situations, but not an issue of general concern (Case Study 1). In the case of the SNSF Fellowship scheme, initiation of the evaluation by actors close to the funding scheme may have resulted in insufficient inclusion of stakeholders advocating more fundamental reform. This may explain why the changes made following the evaluation were soon replaced by further reform (Case Study 11).

The actual evaluation activities seem to be even less of an issue of concern with respect to evaluation use. Methodological quality is quoted as a facilitating factor in three cases (Case Studies 5, 6, and 7). As with evaluator expertise, it may be that informants saw this as a necessary but insufficient condition for evaluation use. The Evaluation of Biodiversity Research in Sweden highlights that even exceptionally highly regarded methodology is not a sufficient condition for use (Case Study 8).

Evaluation planning and design, too, are referred to only on a few occasions. This could be related, in part, to our sample of case studies. Evaluation activities are well established within the organizations we considered and some of the evaluations relied on a proven methodology that was used in earlier evaluations. The experience with HCERES evaluations of INRA indicates that use is facilitated if the emphasis is on future projects of the research units evaluated rather than on achievement, and that summary grades and quantitative indicators are less conducive to use than qualitative explanatory criteria (Case Study 1). With the evaluation of Education Programmes in Norway, the scope aiming to address both scientific and practical relevance of the research funded proved too broad to provide useful recommendations on both aspects (Case Study 9). During the expert validation workshop, too, evaluation of societal impact was considered one of the less mature areas of evaluation.

Evaluation outputs are again more relevant to use, especially timeliness of results, which once more links back to context. In the case of the evaluations of the Emmy Noether Programme and of Biodiversity Research in Sweden, the recommendations lost relevance because they came after important decisions had already been taken (Case Studies 7 and 8). With the evaluations of Plant and Soil Sciences in Estonia and of Mechanical Engineering in Sweden, the results came just in time for important decisions (Case Studies 2 and 3).

The presentation and communication of results, too, was perceived to have an influence on evaluation use in a number of case studies. Detailed recommendations for individual research units or sub-areas of research are perceived to facilitate use with evaluations of research fields and institutes (Case Studies 1–5). Especially recommendations on structural issues were highlighted as prone to be implemented in relation to the evaluation of Mathematics in Sweden (Case Study 3). The evaluations by HCERES, too, were noted to have significant structural implications (Case Study 1).

The case studies underline the importance of evaluation inputs, activities, and outputs. Nevertheless, the overall picture emerges that these are necessary, but not sufficient, conditions to evaluation use. Contextual factors in the end determine whether evaluation studies will be used (in any way).

Overall, the case studies highlight many more factors facilitating use rather than detrimental to use. This suggests that contemporary

evaluation initiatives are generally of good quality. It indicates that the practice of evaluation of research has in many ways evolved into a mature field. Evaluation use, however, depends on multiple, inter-related factors, which need to be considered on a case-by-case basis. The few factors quoted as detrimental to use—budgetary constraints, poor stakeholder consultation, and poor timing of the report—show that the context and the timing of evaluations is a major factor in the use of evaluation studies.

2.3 Who influences the use of evaluation studies

Although it was not explicitly asked in the case study template, in several case studies, the actors involved in the process were mentioned and their role and motives discussed. The following conclusions therefore rely not on a systematic analysis but on the occasional references to actors in the case studies.

Evaluators are—of course—essential in the evaluation process. Their competence, reputation, and engagement determine the credibility and usefulness of an evaluation study.²⁹ However, the importance of context, beyond the immanent evaluation activities themselves, suggests that this is not sufficient and that evaluators have very limited direct influence on the actual evaluation use.³⁰

Evaluations are often expected to fulfil legitimizing, learning, and steering uses in parallel. Especially steering use depends on a constellation of multiple, interacting factors, which are subject to change. The scope for evaluation use needs to be considered on a case-by-case basis, and the political will to implement evaluation results is key. The key to the successful use of evaluation activities may then lie in the interplay between the expertise held by evaluators and the commissioning instances' possibilities to shape the uptake and implementation of recommendations.

It is therefore not surprising that correspondence with established information needs and strategic orientation is crucial, and that the close interaction with stakeholders (e.g. commissioning parties, evaluated institutions, and their research staff) can have strong influence on evaluation use. To involve stakeholders, i.e. to consult with researchers in the programme, field or institute that is being evaluated is often seen as a factor promoting the legitimacy and accuracy of evaluations. They have both in-depth knowledge of the evaluation object and can, to some extent, influence the use of evaluation studies. At the same time, their involvement could distort the evaluation results in favour of these same groups, and thereby threaten the independence of the evaluation.

In our case studies, we did not observe any evidence for the latter. Stakeholder involvement did not appear to be detrimental to the independence of evaluations. Several case studies reported controversial findings such as gender bias or poor strategic planning and human resources management (Case Studies 7 and 5). The evaluation of Plant and Soil Sciences in Estonia as well as certain evaluations by HCERES even resulted in the closure or restructuring of research units/institutes (Case Studies 5 and 1).

Nevertheless, expectations and motivations can vary between stakeholders and unbalanced stakeholder involvement in evaluations may serve existing interest groups and increase the risk of misuse of evaluation for predetermined purposes. To avoid selective, strategic use of evaluation results, it is important to be open about potential conflicts of interest and to clearly and transparently define expectations from the evaluation—which should be oriented

towards original goals of institutions or programmes—as well as responsibilities for follow-up.

3. Conclusions

Expectations of the actual or preferred use of evaluations differ. Evaluators hope that their evaluation study will be used to trigger change (steering use), sceptics feel that evaluations are merely ritualistic or tactical exercises, and funders may feel that evaluations can legitimize a funding scheme. Drawing from a range of evaluation studies in Europe, covering research fields, research institutes and funding instruments, and from a practitioner's perspective, the article has shown that many uses of evaluation studies can (co-)exist. As suggested by Carol Weiss, direct effects on decision-making (steering use) should not be expected from all evaluation studies. Often, evaluation studies contribute to the knowledge base and cognitive resources within organizations and beyond. They may be influential even if their direct use is limited.³¹ Therefore, we suggest reframing the debate and talking about 'influence' rather than 'use'.

Many previous studies and 'how-to' guides have highlighted the need for thorough, high-quality evaluations if they are to have an impact.³² These preconditions remain unquestioned. A flawed evaluation study will not be used or not be used appropriately. A high-quality evaluation study is a necessary, but not sufficient condition to use evaluation results. The study draws attention to the fact that, as stressed for example by Carol Weiss, Claudio Radelli, Bruno Dente, Lorna Schrefler, or—more recently—Jürgen Streicher, the context and other factors external to the evaluation itself, but often in the realm of the commissioning organization, determine the actual use and influence of the evaluation.³³ As evaluation activities are institutionally embedded, their relevance to existing information needs and strategic orientation remains crucial.

The purpose of this study was to test concepts of evaluation use, specifically with respect to the evaluation of research, in various European countries and institutions and concerning different evaluation objects. Overall, the analysis of the case studies has confirmed the findings of many previous studies. Future research could dig deeper into the mechanisms in place, e.g. the motives and specific constellations of actors within research funding or research performing organizations that influence the (types of) use. Furthermore, the impact of codified procedures within commissioning organizations to deal with evaluations deserves a closer look. While many practitioners point to the beneficial effects of a specified follow-up procedure,³⁴ the underlying mechanisms deserve further inquiry. Questions of interest include: How does having a pre-defined management response process or a reflection team accompanying the evaluation influence the way evaluation studies are used? In which way does the hierarchical level at which an evaluation is ordered or the position of the staff included in the discussion process determine the commitment to the recommendations?

From a practitioners' perspective, academic and evaluative thinking about evaluation use can be helpful to inform policies and procedures. Since evaluation uses are multiple and depend on a number of interrelated factors, we would like to suggest the application of logic models to the use of the evaluation itself. Evaluators apply logic models to programmes as a basis for their evaluation activities and could expand the application of these models to the evaluation exercise itself, to consider the underlying mechanisms

and chains of influence through which evaluation may have its effects.³⁵ This could help to target evaluations at knowledge gaps and time them in accordance with policy needs. Also, keeping in mind the different types of use and their interplay can help improving the preconditions for the use of evaluation results. For example, it can point to the importance not only of organizing the discussion of evaluation results to facilitate policy learning but also of disseminating them and creating a wider forum for discussion on the methods and results to serve the information function.

Notes

1. SNSF, DFG, Swedish Research Council (VR), Austrian Science Fund (FWF). Katrin Milzow. Email: katrin.milzow@snf.ch. The article builds on work carried out in a Science Europe Working Group on Research Policy and Programme Evaluation and was written by Katrin Milzow (SNSF Switzerland), Sten Söderberg (VR Sweden), Anke Reinhardt (DFG Germany), and Klaus Zinöcker (FWF Austria). The authors would like to thank Stig Slipersæter (NFR Norway), Marja-Liisa Kärp (ETAG Estonia), Beverly Sherbon, and Ellen Charman (both MRC UK) for their contribution of case studies, as well as all interview partners from evaluating and evaluated organizations. We are grateful to all members of the Science Europe Working Group on Research Policy and Programme Evaluation, to Stéphanie Würth, Michael Hill, and Romaine Brunner (all three SNSF), to Dorothea Sturn, Pierre Glaudes, and Lars Kloo (who participated in the expert discussion), as well as to Gary Henry, for their comments on the report.
2. The findings of this study were validated in a discussion with experts representing different viewpoints on evaluation activities during a workshop in Vienna in November 2016. The experts were Dr Dorothea Sturn, former Managing Director of the Austrian Science Fund (FWF); Professor Pierre Glaudes, Director of the Evaluation of Research Units Department, Haut Conseil de l'Évaluation de la Recherche et de l'Enseignement Supérieur (HCERES); and Professor Lars Kloo, Secretary General, Natural and Engineering Sciences, Swedish Research Council (VR). The authors of this article are grateful to the experts for their valuable contributions. The responsibility for the statements and any mistakes of course lies with the authors.
3. See amongst others, Nutley et al. (2010); Biegelbauer 2013; Straßheim and Kettunen 2014.
4. European Science Foundation 2012a
5. Scriven 1991; Patton 1997; Streicher (2017).
6. Science Europe is a collaborative association of European Research Funding Organizations and Research Performing Organizations with 43 Member Organizations from 27 countries.
7. In particular Streicher (2017) op. cit.; Mark and Henry (2004); Vedung (1997).
8. As suggested by Streicher (2017) op.cit, 157.
9. Vedung (1997) op. cit.
10. Fleischer and Christie (2009).
11. Weiss (1979).
12. Weiss (1979) op. cit., 426–30.

13. Schrefler (2010); Boswell (2008).
14. Platform Research and Technology Policy Evaluation (fteval) (2005).
15. This can reach from a basic provision of data and information to the deliberate use of evaluation as a source of 'strategic intelligence'; see Kuhlmann (2003).
16. Vedung (1997), op. cit.; see also Forss, Rebien and Carlsson (2002).
17. Also see Weiss (1979) op. cit., 429.
18. The term 'ritual' is often negatively connotated and in this context also used to criticize evaluations that do not seem to have a learning purpose. See Schwarz (2006); Floden and Weiner (1978).
19. Johnson et al. (2009); Fleischer and Christie (2009) op. cit.; Streicher (2017) op. cit.
20. Weiss (1998); Radaelli and Dente (1996); Schrefler (2010) op.cit.
21. Radaelli and Dente (1996) op. cit., 60.
22. Mark and Henry (2004).
23. Sager and Pleger, 2016; OECD 1991, 5.
24. Evaluation objects adapted from European Science Foundation 2009, 5.
25. Seawright and Gerring (2008).
26. See for example the findings of Streicher (2017) op. cit., 4.
27. Mark and Henry (2004) op. cit., 35, 37.
28. This echoes the results of Streicher (2017) op. cit., 4.
29. See Patton et al. (1977).
30. See also Weiss (1998) op. cit., 2–23.
31. Weiss (1979) op. cit., 428.
32. The conclusions that we draw build on previous work by the European Science Foundation (2012b) op. cit., which stressed that potential uses should be considered from the start when planning evaluations, and that sufficient time and resources should be reserved for follow-up. The ESF also generally stressed the importance of timeliness, methodological quality, and stakeholder involvement.
33. Weiss (1998) op. cit., 23; Radaelli and Dente (1996) op. cit.; Schrefler (2010) op.cit.
34. See European Foundation Centre (2010); European Science Foundation (2012b); Reinhardt (2013).
35. See also Mark and Henry (2004) op. cit., 35.
36. Source: Appendix to our case study template; First five types: FTeval, Research and Technology Policy Evaluation, op. cit., Tactical use: Vedung (1997) op. cit., Process use: Forss, Rebien and Carlsson (2002) op. cit.

Conflict of interest statement: The authors are all employed by research funding organizations that contributed case studies. The case studies and analytical work were performed as part of the authors' employment and within the framework of a collaborative project of Science Europe.

References

- Biegelbauer, P. (2013) *Wie Lernt Die Politik?: Lernen Aus Erfahrung in Politik Und Verwaltung*. Wiesbaden: Springer-Verlag.
- Boswell, C. (2008) 'The Political Functions of Expert Knowledge: Knowledge and Legitimation in European Union Immigration Policy', *Journal of European Public Policy*, 15/4: 471–88.

- European Foundation Centre. (2010) *Research Forum Evaluation Guidelines 3: How to Evaluate*. Brussels: European Foundation Centre.
- European Science Foundation. (2009) *Evaluation in National Research Funding Agencies: Approaches, Experiences and Case Studies*. Strasbourg: European Science Foundation.
- European Science Foundation. (2012a) *Evaluation in Research and Research Funding Organisations: European Practices*. Strasbourg: European Science Foundation.
- European Science Foundation. (2012b) *A guide to Evaluation Activities in Funding Agencies*. Strasbourg: European Science Foundation.
- Fleischer, D. N., and Christie, A. C. (2009) 'Evaluation Use: Results from a Survey of U.S. American Evaluation Association Members', *American Journal of Evaluation*, 30(2): 158–75.
- Floden, R. E., and Weiner, S. S. (1978) 'Rationality to Ritual: The Multiple Roles of Evaluation in Governmental Processes', *Policy Sciences*, 9(1): 9–18.
- Forss, K., Rebienc. C., and Carlsson J., (2002) 'Process Use of Evaluations, Types of Use That Precede Lessons Learned and Feedback', *Evaluation*, 8(1): 29–45.
- Platform Research and Technology Policy Evaluation (fteval) (2005), 'Evaluation Standards in Research and Technology Policy'. <<https://wbc-rti.info/object/document/7019/attach/EvaluationStandardsinRTDfteval.pdf>> accessed 30 Oct 2017.
- Johnson, K. et al. (2009) 'Research on Evaluation Use: A Review of the Empirical Literature from 1986 to 2005', *American Journal of Evaluation*, 30(3): 377–410.
- Kuhlmann, S. (2003) 'Evaluation as a Source Of strategic Intelligence', in: Shapira P. and Kuhlmann S. (eds) *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*, pp. 352–79. Cheltenham: Edward Elgar Publishing.
- Mark, M. M., and Henry, G. T. (2004) 'The Mechanisms and Outcomes of Evaluation Influence', *Evaluation*, 10: 35–7.
- Nutley, S. et al. (2010) 'Evidence and Policy in Six European Countries: Diverse Approaches and Common Challenges', *Evidence and Policy*, 6(2): 131–44.
- OECD (1991) *Principles for the Evaluation of Development Assistance*. Paris: OECD.
- Patton, M. Q. et al. (1977) 'In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research', in Weiss C. (ed.) *Using Social Research in Public Policy-Making*, pp. 141–63. Lexington, MA: Lexington Books.
- Patton, M. Q. (1997) *Utilization-Focused Evaluation: The New Century Text*. London: Sage.
- Radaelli, C. M., and Dente, B. (1996) 'Evaluation Strategies and Analysis of the Policy Process', *Evaluation*, 2(1): 51–66.
- Reinhardt, A. (2013) 'Das Evaluationsvorgehen Der Deutschen Forschungsgemeinschaft', *Zeitschrift Für Evaluation*, 2: 305–14.
- Sager, F., and Pleger, L. (2016) 'Evaluation and Independence, Existing evaluation policies and new approaches', Report for the UNDP independent evaluation office.
- Schrefler, L. (2010) 'The Usage of Scientific Knowledge by Independent Regulatory Agencies', *Governance: An International Journal of Policy*, 23(2): 309–30.
- Schwarz, C. (2006) *Evaluation als Modernes Ritual: Zur Ambivalenz Gesellschaftlicher Rationalisierung am Beispiel Virtueller Universitätsprojekte*. Hamburg: LIT Verlag Münster.
- Scriven, M. (1991) *Evaluation Thesaurus*. Newbury Park: Sage.
- Seawright, J., and Gerring, J. (2008) 'Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options', *Political Research Quarterly*, 61(2): 294–308.
- Straßheim, H., and Kettunen, P. (2014) 'When Does Evidence-Based Policy Turn into Policy-Based Evidence? Configurations, Contexts and Mechanisms', *Evidence and Policy*, 10(2): 259–77.
- Streicher, J. (2017) 'Evaluations, Actors and Institutions. The Case of Research, Technology and Innovation Research in Austria', Doctoral Thesis, WU Vienna University of Economics and Business.
- Vedung, E. (1997) *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction.
- Weiss, C. (1979) 'The Many Meanings of Research Utilization', *Public Administration Review*, 39(5): 426–31.
- Weiss, C. (1998) 'Have We Learned Anything New about the Use of Evaluation?', *American Journal of Evaluation*, 19(1): 21–33.

Appendix 1: Case studies

Case Study 1: Evaluation of research units of the Institut national de recherche Agronomique (INRA) by the Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES).

Research units in France are evaluated every 5 years by HCERES in a methodologically standardized fashion. The evaluations include qualitative as well as quantitative aspects. They aim, for example, to highlight the strengths of a particular site, trying to describe why the characteristic is perceived as a strength. The evaluations result in recommendations to amend a unit's area of activity, to continue a unit or not or to merge units. Explicit recommendations of this kind facilitate the use of the evaluations. At INRA, the evaluations have had learning effects for the units and departments concerned. They have affected budget decisions and led to structural change in French research towards the general objective of reducing the number of entities and establishing more interdisciplinary groups.

Case Study 2: Swedish Research in Mechanical Engineering (VR)

The evaluation can be regarded as a traditional evaluation of a research field, both in terms of methodology (informed peer review) and timing (fields are evaluated on a regular basis). At the same time, there was a concern about a possible decline in the level of quality in the research field of Mechanical Engineering in Sweden and some considerations about reducing the budget for funding applications at the VR for this particular research field.

The aim was to evaluate all research in Mechanical Engineering in Sweden, regardless of funding. Based on self-assessments and bibliometric information, the evaluation panel reported detailed assessment and conclusions for six sub-areas of research in Mechanical Engineering and each institution concerned. The panel also gave some brief general recommendations for the research field of Mechanical Engineering in Sweden as a whole, mainly consisting of a SWOT analysis on a national level.

The main effect of the evaluation was that VR's Scientific Council for Natural and Engineering sciences made no decision to reduce the budget for applications for the research field of Mechanical Engineering. Thus, the possible 'threat' to the funding level of VRs support for research in Mechanical engineering was abated.

Case Study 3: Swedish Research in Mathematics

Two Swedish funders commissioned the evaluation, which addressed a knowledge gap regarding the evolution of the research field since the prior evaluation ca. 15 years earlier and concern about a decline in the level of quality in the research field. The evaluation was conducted by an international expert committee. Swedish mathematicians and heads of Mathematics departments were able to suggest members. Over 100 project leaders were asked to write self-assessments, and there were hearings with the heads of the concerned departments (at 13 Swedish universities).

The decision to restrict the selection of research projects to those projects funded by the two funders was criticized, leading the VR to adjust the evaluation design in its forthcoming research field evaluation of Mechanical Engineering. In this latter evaluation, the design was based on a statistical (register-based) analysis of the research field in question.

In the Mathematics research community, there was a perception that the field was under-funded and 'below the radar'. The community welcomed the evaluation, since there was a perception that a decision to evaluate a research field automatically puts the spotlight on the field guaranteeing attention. The subject of Mathematics was widely discussed in Swedish educational policy debates (e.g. the TIMSS and PISA results), so there was a good breeding ground for initiatives.

Detailed conclusions and recommendations for both sub-areas of research and for individual institutions, many of them structural, were perceived as very useful to VR and other funders. The teaching-intense nature of the subject, in combination with a limited number of funding sources, had led to a situation where funding of research in Mathematics was low and ineffectively used. The results were seen as an alarm clock—'something needs to be done' and the evaluation resulted in a well funded new 'Mathematics initiative'.

Case Study 4: Research in Mathematics at Norwegian Universities

The Norwegian government periodically commissions subject-specific evaluations. These evaluations therefore have a 'ritual' aspect. The evaluations are typically associated with a sum of money for strategic projects. Participating institutions able to suggest international experts for the high-level evaluation committee saw the evaluation as an opportunity to direct attention to their field and get more funding for research in Mathematics. However, some institutions chose not to participate. Institutions did not see the results of the evaluation before they were published but were called on to describe their follow-up plans and suggest actions at the national level. This feedback was then processed by a 'follow-up committee' with representatives from research institutions. The adoption of recommendations from prior evaluations formed part of the evaluation questions.

The evaluation provided information on the high scientific quality of the field and thus also served legitimizing purposes. It contributed to the development of the Research Council of Norway's main funding scheme, notably the introduction of funding opportunities to further define career structures and tenure track positions, development of world-leading research groups, and increased international mobility among postdoctoral fellows. The funds for institutional strategic projects went to a research school in bioinformatics.

Case Study 5: Plant and Soil Sciences in Estonia

The evaluation was initiated by the Estonian government (Ministry of Education and Research and Ministry of Agriculture). It addressed concerns about research quality and responded to a real need and demand for information from the government as a basis for restructuring the research field. The government therefore had a clear interest in designing an evaluation that would give practical results and in using the results. The evaluation commission was very high level and very respected. The evaluation was widely discussed with different stakeholders. The tailor-made design was confirmed by representatives from research fields and ministries. Stakeholders also expected that the evaluation would lead to a restructuring and did not contest the proposed changes. The evaluation led to concrete decisions to restructure two research institutes, and overall to strengthened efforts for more co-operation and mobility.

Case Study 6: DFG's Collaborative Research Centres—Transfer Projects

The evaluation focussed on the functioning of the funding scheme to support knowledge transfer. It was commissioned by the DFG and executed externally. The evaluation team was specifically asked to focus on collecting information but not to give recommendations. This was the task of a steering committee that was established for this evaluation. The committee was satisfied with the report. It found no need for programme adaptation, although a call for more 'marketing' of the programme was made. Budget cuts hindered the implementation of some recommendations.

Case Study 7: Deutsche Forschungsgemeinschaft's Emmy Noether Career Programme

Several years after inception, the Emmy Noether programme was to be evaluated. The objectives were to evaluate the effectiveness of the Emmy Noether programme and its relation to applicants' career prospects, their situation as junior group leaders, and their work-life balance, etc. The overall conclusion was that the programme delivers the expected results as a highly successful career scheme. Nevertheless, the evaluation did report some negative findings, such as a small gender bias. Overall, no need for programme adaptation was indicated according to the joint committee, the highest decision-making body of the DFG. Some changes to the programme, which were also called for in the recommendations of the evaluation, were already implemented, however, either before publication of the study or shortly after. Nevertheless, the publication of the study generated substantial interest and impact: already shortly after publication, it had been downloaded more than 1,500 times.

Case Study 8: Swedish Research in Biodiversity

This evaluation responded to information needs about the consequences of a new political programme to encourage a specific research field. The evaluation methodology was considered excellent by several stakeholders and subsequently influenced several other evaluation studies, which adopted the idea of a separate relevance committee. It was used mainly for legitimizing purposes, to show that extra funding had been used to fund research of high quality and relevance, thus dismissing the concerns of some critics. The evaluation's main recommendation—to continue earmarked support for biodiversity research—was followed by neither of the two organizations concerned, the VR and Formas. In fact, following a change in government, the earmarked funding for biodiversity research was in the meantime abolished, and VR had decided already during the course of the evaluation to suspend special management arrangements for the funding of biodiversity research. Some steering use can be observed at Formas, where the evaluation findings complemented several other data sources underlying a decision to abolish earmarked support to Postdocs and Assistant Professors for all research fields.

Case Study 9: Education programmes in Norway

This integrated evaluation of two partially overlapping programmes—The Programme for Educational Research (UTDANNING2020) and the Programme for Practice-based R&D in Pre-school through Secondary Schools and Teacher Education (PRAKSISFOU)—was called for by the Research Council of Norway and commissioned by the Norwegian government. It was conducted by a panel of experts from Sweden, Denmark, and Norway. It provided a knowledge base for and served to legitimate the fusion of the two programmes. The evaluation also helped to mediate between different stakeholders. The evaluation report was of good quality and well received, although it was felt that one aspect of the spectrum of evaluation questions (practical impact of the programme) was neglected.

Case Study 10: National Prevention Research Initiative (NPRI, UK)

The 16 NPRI funders (including government departments, research councils, and medical charities) commissioned the evaluation in February 2014 to assess the individual project outputs and the initiative's strategic impact. The evaluation was also commissioned to advise on future opportunities in prevention research. The Scientific Review Group in charge of the evaluation found that few interventions had produced large practical change and that the changes that did occur were mainly produced by researchers with a history of interaction with research users. The implementation of recommendations is perceived to depend above all on budgetary decisions.

Case Study 11: Swiss National Science Foundation (SNSF) Fellowship Schemes

The evaluation of SNSF Fellowships for Early and Advanced Researchers was commissioned in 2009 on the impulse of the SNSF's Specialized Committee for Careers. The main motivation was questions about operational issues regarding the configuration of the grants (grant size, grant duration, employment situation, and social security for fellowship holders) and the evaluation procedure (evaluation in part by local research commissions at the universities, postdocs, and PhDs in parallel). The evaluation focused mainly on the fellowship schemes and their attractiveness to applicants and to some extent on their link to other SNSF funding instruments. As a result, a new funding scheme was created. The career funding schemes are however currently subject to further reform.

Appendix 2: Case study template**General questions:**

1. How were evaluation activities generally integrated into the governance of the organization evaluated at the time of the evaluation? Was there anything specific/different about this evaluation?
2. Who commissioned the evaluation? What were the motives for conducting an evaluation?
3. What were the objectives of the evaluation?
4. Who conducted the evaluation?
5. Who was involved in the design of the evaluation? Were stakeholders involved in the course of the evaluation? Who and how? Were there players in favor of changing programmes/procedures? Players opposing change? Was there anything particular about the context of the evaluation?
6. Brief description of methods, findings, conclusions, and recommendations

Specific questions on evaluation use:

7. What were the intended uses of the evaluation?
8. Was there anything in particular (perhaps also unintentional or unexpected) that influenced the evaluation use? Were there any

particular challenges preventing the evaluation from having an impact or reducing its relevance? Any factors facilitating evaluation use?

9. Description of actual use(s):
 - a. Explanation of the process or means through which the evaluation (process or result) came to have an influence: How did the evaluation play a role, and how were its results taken up or applied?
 - b. Details of the users/stakeholders: Who/what constituency/what organization has been affected?
 - c. How have they been affected?
10. Which type(s) of use were relevant to this evaluation? The proposed categorization (see Chapter 2.1) may serve as a basis, but there may be other types of use which should also be considered.
11. Where possible, evidence of use—e.g. references in reports or web contents, statements by stakeholders. . .
12. Was there discussion about the quality of the evaluation or evaluation report?

Open questions:

13. Is there any additional information relating to the evaluation, its context, or its use that might be relevant to the analysis?

Appendix 3: Template for case study analysis

Date of publication of evaluation report: 20XX

1. Which of the following types of evaluation use can be identified in this case study? How relevant are they?³⁶

	Not at all				Very much
Legitimation For example, justifying the use of public funds	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Information providing the public with information on how public funds are being used and to what effect	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Learning for funding institutions, decision-makers in the field of technology policy, intermediaries, universities, research institutions, enterprises, consultants, and scientists	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Steering for establishing policy objectives and planning of measures	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mediating between the competing interests of different players	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tactical to gain time or avoid responsibility	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ritual Systematic evaluation at certain intervals/on certain occasions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. Which factors facilitated the influence/use of the evaluation?

	Short explanation where applicable
Context For example, institutional embeddedness of evaluation activities, information need/availability, Evaluands' culture, expertise and impetus, and macro- and micro politics	<ul style="list-style-type: none"> • Example of use for institutional embeddedness: generally a certain sum of additional money available is made available to implement recommendations from evaluations • Example of use for culture: evaluands were used to evaluation, open to suggestions from evaluators
Input For example, instruction, communication, time, resources, information/data, evaluators' expertise, and reputation	<ul style="list-style-type: none"> • Example of use for expertise: high reputation of evaluation panel furthered implementation of findings
Activity For example, stakeholder selection and participation, evaluation planning and design, methodological planning, data collection and analysis, developing conclusions and recommendations, and report generation NB: Process use can be covered under this heading	<ul style="list-style-type: none"> • Example of use for evaluation planning and design: Joint evaluation by different stakeholders increased acceptance and facilitated implementation of recommendations
Output For example, form of report, presentation/communication of results, perceived methodological quality/credibility, and timeliness	<ul style="list-style-type: none"> • Counterexample of use for methodological quality: evaluation methodology was considered best practice, but no use of evaluation can be identified • Example of influence for methodological quality: evaluation methodology considered best practice, was taken up for further evaluation studies

3. Which factors were detrimental to the influence/use of the evaluation?

	Short explanation where applicable
Context For example, institutional embeddedness of evaluation activities, information need/availability, Evaluands' culture, expertise and impetus, and macro- and micro politics	-
Input For example, instruction, communication, time, resources, information/data, evaluators' expertise, and reputation	-
Activity For example, stakeholder selection and participation, evaluation planning and design, methodological planning, data collection and analysis, developing conclusions and recommendations, and report generation	-
Output For example, form of report, presentation/communication of results, perceived methodological quality/credibility, and timeliness	<ul style="list-style-type: none"> • Example for quality: concern about quality used as a pretext to avoid engaging with findings • Example for timeliness: lengthy process conflicted with impulse for change, and changes were made before the evaluation report was published • Example for communication: disrespect of lines of reporting harmed acceptance of evaluation findings and hindered use of results

4. Does the case study support more general findings like the following? Can other, similar statements be identified, which can be supported by the case study?

	Yes/no, comments
Evaluations are more likely to be used when particular knowledge gaps are addressed	
Changes are made before the evaluation report is published	
Evaluation reports are always positive about evaluation objects	
There may be goal conflicts between stakeholder involvement and independence	
Methodological quality does not guarantee that evaluation results are used	
The evaluation had an influence beyond the evaluation object	