Power in and of evaluation: A framework of analysis

Anders Hanberger
Umeå University, Sweden

Abstract
This article develops a framework for enhancing understanding and exploring both how power manifests in the evaluation process, and the power of evaluation in relation to public policy and democratic governance. Power is conceived as a multifaceted and dynamic phenomenon that manifests, permeates, and affects evaluation in many ways. The article demonstrates how the framework can be applied to an evaluation of a Swedish teacher-training program. The tentative analysis shows how the commissioner’s power-over the evaluators becomes evident when it cannot induce the evaluators to do what it wants them to do and manifests itself as constitutive power when, for example, helping shape the notion of what valid knowledge is. The power of the evaluation manifests itself as supporting key policy and governance functions.

Keywords
democratic governance, evaluation, framework, power, public policy

Introduction
Who has and should have power in evaluation and what is and should be the role and power of evaluation in public policy and governance are basic questions that need recurrent discussion, particularly when the conditions for evaluation are changing. In the last few years, the conditions for evaluation have changed and created new challenges for evaluation. The changes can be understood and described in different ways. Dahler-Larsen (2021) describes them as four shifts away from research-based stand-alone evaluations that inform “political or managerial decision-making” in representative democracy. There has been a shift from government to governance, from ad hoc evaluation to evaluation systems, a shift in epistemic foci (with less emphasis on how a policy works and more emphasis on performance criteria) and a shift in the status of validity and reliability (“from being seen as logical preconditions of...
evaluative information to being contextual factors of variable importance”) (Dahler-Larsen, 2021: 20). Evaluation has become more institutionalized and released from democratic control (Dahler-Larsen, 2021). These and other changes in the conditions for evaluation (Hanberger, 2012, 2018; Picciotto, 2015) imply a need for further discussion and exploration of power and evaluation in public policy and democratic governance.2

The article contributes to knowledge of power and evaluation by developing a framework for enhancing understanding and exploring both how power manifests and evolves in the evaluation process, and the power of evaluation in public policy and democratic governance. It also contributes to knowledge of how power in and of evaluation relates in this context.

Largely, power in evaluation and power of evaluation have been treated as separate issues. However, the relation between the two issues is evident in the context of public policy and governance. The relation can manifest in different ways in different governance settings. Generally, a relation between the “in” and “of” manifests when viewed through the lens of constitutive power (Haugaard, 2010) and as constitutive effects of evaluation (Dahler-Larsen, 2012, 2014). Within a governance model, an institutionalized evaluation (system) constitutes a shared epistemological perspective (Leeuw and Furubo, 2008) and shapes the language and social interactions between actors (Andersen, 2020). However, how power manifests and evolves in different evaluations and governance settings are empirical questions. How actors use their delegated or mobilized power (Arendt, 1970; Barnes, 1988; Dahl, 1957; Weber, 1978) to influence evaluation (systems) in the evaluation process can affect the design and implementation of evaluation, the evaluation reports and the potential and actual power of evaluation in public policy and governance.

Against this background, the article further discusses and explores power and evaluation from a democratic governance perspective, how power manifests and evolves in the evaluation and policy process through actors’ mobilization of power and when it manifests itself as constitutive power. It contributes to enhancing understanding and knowledge of why actors act the way they do in the evaluation process and how their action together with constitutive effects of evaluation affect evaluation reports and the power of evaluation to support key functions in democratic governance.

To gain comprehensive understanding and knowledge of power and evaluation in this context, the article integrates notions of power through structure and agency. It conceives power as a multifaceted and dynamic phenomenon that manifests and affects evaluation in many complementary and sometimes conflicting ways.

The aim is to develop a framework to enhance understanding and support exploration of power in and of evaluation from a governance perspective. Answers are sought to the following two related questions:

1. How can power be understood and explored in different phases of the evaluation process?
2. How can the power of evaluation be understood and explored in relation to key functions in policy and governance?

The article continues with a brief overview of how power is discussed in the political–philosophical literature and the evaluation literature. The article then, based on this literature and research on evaluation and governance, develops the framework. Next, it demonstrates how the framework can be applied to the evaluation of a Swedish teacher-training program. It
ends with conclusions about the knowledge generated by the framework and discussion about its advantages and limitations and needs for further research.

**Notions of power in the political–philosophical literature**

Evaluation researchers borrow, explicitly or implicitly, concepts of power from the political–philosophical literature. Therefore, the article begins with a brief overview of how power is conceived and discussed in this literature in relation to decision-making.

This large and growing literature is characterized by deep disagreements over how power should be understood and defined. The discussion of the nature of power is ongoing, reflecting that power is a multifaceted, contested, and dynamic phenomenon.

One main disagreement is between those who understand and define power as getting someone else to do what you want them to do (referred to as “power-over”) and those who define it as the ability or capacity to act (referred to as “power-to”). Classical definitions of the former are Max Weber’s (1978) definition of power as “the probability that one actor within a social relationship will be in a position to carry out his own will despite resistance” (p. 53) and Robert Dahl’s (1957) definition “A has power over B to the extent that he can get B to do something that B would not otherwise do” (pp. 202–203).

Hannah Arendt (1970) defined power as “the human ability not just to act but to act in concert” (p. 44), which is a classic definition of the power-to notion. Amy Allen (1999) extended Arendt’s notion by defining power as “the ability of a collectivity to act together for the attainment of an agreed-upon end or series of ends” (p. 127). However, and as Steven Lukes (2005) recognized, power is a capacity that may or may not be used; he maintains that power “is a potentiality, not an actuality—indeed a potentiality that may never be actualized” (p. 69). These notions of power can be used to reflect actors’ use of power in the evaluation and policy process, but there are more notions that need attention in decision-making. According to Rye (2015), an eclectic approach provides a richer and more flexible approach to studying power in organizational contexts.

Bachrach and Baratz (1962) argued that power is exercised not only through decision-making itself (first face), but also by non-decision-making (second face), that is, by excluding issues from the political agenda. Lukes (2005) recognized a third face of power, with less focus on actors’ behavior and more on the hidden and subtle ways power manifests. Lukes argued that power evolves both through actors’ behavior and through structures, and that power favors certain interests over others. The “dominated” are often unaware of the subtle ways power operates. The fourth face of power, based on Foucault’s (1977) notion of disciplinary power, reflects the fact that societies produce modern and postmodern subjects and identities and is a form of power that is implicit, sometimes invisible and shaped through discourses, knowledge production, governance techniques, and institutional arrangements (Digeser, 1992). In addition, the systemic or constitutive conceptions, according to Haugaard (2010), view power as “the ways in which given social systems confer differentials of dispositional power on agents, thus structuring their possibilities for action” (p. 425). This notion takes into account “the ways in which broad historical, political, economic, cultural, and social forces enable some individuals to exercise power over others, or inculcate certain abilities and dispositions in some actors but not in others” (Allen, 2016: n.p.). The constitutive and disciplinary notions of power can be used to reflect the way power manifests through an evaluation system, for example.
Power has more connotations as it is often combined with other terms. Empowerment is one term frequently used and is of interest for this article as evaluation can empower or disempower actors in governance. One kind of empowerment comes from “downward delegation,” defined as when “an existing power-holder delegates some of his/her capacity for action to a subordinate” (Barnes, 1988: 71). When an agent is empowered by a power-holder, she is given discretion to decide and act under certain conditions regarding certain matters. In contrast, Wartenberg (1990: 207) discussed power that is beneficial to both parties, using the term “transformative power” when the relationship between power-holder and subordinate is more likely to be mutually beneficial (in ideal cases). It is a voluntary relationship that requires openness and trust. Most notions view empowerment as a zero-sum game, overlooking the fact that empowerment is largely a collective resource that can grow, as empowering some does not automatically entail the disempowerment of others (Ball, 1992).

As pointed out by Haugaard (2010) and Allen (2016), how we conceptualize power is highly shaped by the political and theoretical interests that we bring to the study of power. It is also recognized that our conceptions of power are themselves shaped by power relations (Allen, 2016; Lukes, 2005). Lukes (2005) underscored that “how we think about power may serve to reproduce and reinforce power structures and relations, or alternatively it may challenge and subvert them” (p. 63). Hence, how evaluation researchers conceive and manage power issues in evaluation can either reinforce established power structures or challenge them.

How evaluation researchers apply these and other notions of power is discussed next.

Notions of power in the evaluation literature

Rutkowski and Sparks (2014) recognize that evaluation occurs in a complex political terrain where international organizations can assume some sovereignty. They conceive power as a resource and as national political power-over international organizations and evaluation. Similarly, Eckhard and Jankauskas (2019) assume a resource-based notion of political power, conceived as power-over and power-to act focusing on how stakeholders can influence evaluation with agenda-setting power and other political resources. Furubo and Karlsson Vestman (2011) discuss the power of evaluation itself, claiming that the evaluator is a stakeholder with “its own interests and power dynamics to safeguard.” Evaluators’ power can manifest as safeguarding future evaluation commissions or defending the chosen evaluation approach, for example. These authors assume an actor-based notion of power viewing power as a resource that power-holders (national politicians and evaluators) can use to influence evaluation.

In contrast, Raimondo (2018) assumes a constitutive notion of power when exploring evaluation systems. She argues that “evaluation systems derive their power from their capacity to structure knowledge, establish and diffuse norms about worthwhile interventions inside and outside organizations” (Raimondo, 2018: 35). This framework assumes three main ways in which international organizations enact their power: “classification,” “meaning-making,” and “diffusion of norms.”

Evaluation (systems) can shape language, norms, what is considered valid knowledge, and social interactions among actors (Andersen, 2020; Dahler-Larsen, 2012; Furubo and Karlsson 2011). Andersen (2020) recognizes that an evaluation system can “redistribute power and authority from politicians, interest-groups and citizens to civil servants with the most analytical capacity” (p. 270).

Some evaluation models and approaches are explicitly developed to manage power imbalances in evaluation. Baur et al. (2010) discuss how asymmetric power relations among
stakeholders can be managed by evaluators in responsive evaluation. Power imbalances can be used constructively by establishing trust, including marginalized groups, and promoting dialogue and mutual learning processes among stakeholders that empower all stakeholders. The aim is not to reach consensus on problems or future actions but to gain respect and understanding:

We argue that it is essential for responsive evaluation that the evaluator does not try immediately to ease conflict or silence powerful voices in a dialogue. Instead, evaluators should work with stakeholders towards a situation in which all feel empowered to work on practical improvements together. It is a shared responsibility of all stakeholders to solve conflicts and to learn to hear silenced voices. The evaluator can help create awareness about this and he holds a mirror up to stakeholders. (Baur et al., 2010: 245)

These authors also assume an eclectic notion of power combining power-over, power-to act, and transformative power.

Haugen and Chouinard’s (2019) conceptual model of power, developed for analyzing power in culturally responsive evaluations (CREs), is of special interest here as it conceptualizes how power can manifest in the evaluation process. Their model, derived from a research synthesis of CREs and based on Foucault, frames power as a dynamic, relational, and productive concept that structures knowledge and that manifests at multiple levels in evaluation settings (Haugen and Chouinard, 2019: 378). The model recognizes that power can be visible, hidden, or invisible.

Relational power prevails between and among all members in the evaluation process. Power relations are socially and politically shaped and influence relationships between evaluators and stakeholders. “Given the nature of most evaluation contexts and the relational dynamics that are inherent among human beings, unspoken or invisible power dynamics are often diffused among and across participants and evaluators and can greatly affect the evaluation process” (Haugen and Chouinard, 2019: 378). Political power is understood as the power that shapes “the barriers and biases that prevent people from participation, influencing which voices and perspectives will be included and which will be excluded” (Haugen and Chouinard, 2019: 379). This notion of power includes “the macrostructures of inequality (Kothari, 2001) around issues of gender, citizenship, sexuality, and class” (Haugen and Chouinard, 2019: 380). Discursive power refers to the power of producing the “truths” that govern our lives and structure “reality” and what is considered valid knowledge. It invisibly shapes beliefs, values, and norms. Historical/temporal power refers to a kind of power that is socially constructed prior to the evaluation, that is, power “as historically and politically mediated by often invisible social and institutional forces that originate outside of the parameters of the local setting” (Haugen and Chouinard, 2019: 381).

This conceptual model provides a multidimensional understanding of how power can manifest in CRE processes. Although the framework recognizes many dimensions of power that may be in play in the evaluation process, the authors do not suggest how to explore these dimensions or how the governance structure (included in political power) constitutes and affects power in the evaluation and policy processes.

Dahler-Larsen’s (2015) discusses what largely affect the power of evaluation. Evaluation is a social practice set up to change another practice (evaluand) and to achieve this it must protect itself from contestability (cf. Haugaard, 2010). The evaluation must be backed up by such as trust in the applied methodology, data and the institution that carries out evaluation and in
“the virtues related to using evaluation for good purposes such as learning or improvement” (Dahler-Larsen 2015: 31). Moreover,

To function effectively, an evaluation must exploit the differential between the (relative) fluidity of the social material it seeks to change and the (relative) solidity of its own fixation in the world. I call this difference “the contestability differential.” All evaluation plays with the difference between what is solid and what is not solid. (Dahler-Larsen 2015: 31)

This implies that if an evaluation is not conceived as solid, its authority and power diminishes. However, this is but one factor that, under certain conditions, can influence evaluation use (Dahler-Larsen, 2021) and the power of evaluation.

Andersen (2021) recognizes that an evaluation system can both increase power-over evaluands and decrease the power of evaluation to promote change. “On the one hand, this [the evaluation system] maximises compliance with the recommendations of evaluations—thus increasing evaluations’ power over evaluands. On the other hand, this fixation of subject-positions and epistemological perspective also decreases evaluations’ power to invoke radical change and development” (Andersen, 2021: 53). His observation indicates a relation between power in and of evaluation and the article will discuss this relation and paradox later.

The cited literature contributes to a many-sided understanding of different dimensions and aspects of power and evaluation. However, it provides limited guidance for exploring power in and of evaluation as related issues in democratic governance. This justifies developing the framework, which this article sets out below.

**Framework**

The framework, illustrated in Figure 1, is based on the cited research on power and research on evaluation and governance (Bovens et al., 2006; Chelimsky, 2009; Dahler-Larsen, 2012, 2015, 2021; Furubo and Karlsson, 2011; Hanberger, 2006, 2009, 2011, 2012; Howlett, 2014; Klijn, 2008; Leeuw and Furubo, 2008; Picciotto, 2015; Schoenefeld and Jordan, 2017; Trochim, 2009; Weiss, 1993, 1999; Widmer and Neuenschwander, 2004). It recognizes that public-sector evaluation is embedded in a governance structure (with a division of power between levels of government) and governance model (e.g. management by objectives) and guided by an evaluation policy. Political and administrative actors use their power to set the rules of the evaluation in line with the governance structure, governance model, and evaluation policy, which frames and governs the evaluation to the functions it should support in governance (Hanberger, 2011, 2012). The framework treats power in and of evaluation as related issues embedded in the same governance structure and governance model.

The framework conceives power as a multifaceted and dynamic phenomenon that permeates and affects evaluation and is enacted through structure and agency. Power can manifest as power-over, power-to act, empowering actors, transformative power, constitutive, and disciplinary power (not illustrated in the figure).

First, as indicated by downward arrows the governance structure, governance model and evaluation policy both affect power in the evaluation process and the power of evaluation to support key functions in democratic governance. For example, Management by Objective requires objectives oriented evaluations aimed to support the policy improvement and accountability functions. This governance model and evaluation model do not (need to) involve stakeholders in the evaluation process. In contrast, network governance needs stakeholder
evaluations to also support the collective learning and democratic discourse functions. Key stakeholders are therefore involved and delegated power in the evaluation process. It is, however, an empirical question how actors use and mobilize power in the evaluation process and if an evaluation (system) actually supports its intended function(s).

The horizontal dotted arrow indicates that there is some kind of relation between power in and power of evaluation in this context. The relation can manifest through constitutive power (see introduction), and if actors in the evaluation process use their power to affect evaluation reports and then refer to the evaluation in the continuing policy process to support/impede key functions in public policy and governance.

Second, the framework directs attention to how the above notions of power can manifest and evolve in the following four phases of the evaluation process (first column in Figure 1): the design phase (I), the implementation phase (II), the conclusion phase (III), and the communication phase (IV). During the design phase (I), the preconditions for the evaluation are interpreted and the evaluation model put into practice, providing roles and power to actors at the outset of the evaluation process. How actors use their power to comment on the design can affect the evaluation reports. In a formative evaluation, the shaping of policy (or program) continues during the implementation phase (II) and actors can use their power to influence the development of the evaluation during its implementation. During this phase, solutions to emerging issues regarding selection of cases, data collection, and unexpected challenges may trigger actors to use their power. In the conclusion phase (III), power dynamics can manifest depending, for example, on how conclusions and recommendations in a draft report are received. Power in the communication phase (IV) often manifests in what the actors choose to highlight, overlook, or misrepresent with reference to the evaluation (system). If findings are disputed, actors can use
their power to question, for example, the evaluation design, the evaluator’s competence, or the conclusions’ empirical support. At this stage, the evaluation process merges with the policy process (Hanberger, 2011). Whether and how actors in the evaluation process and the continuing policy process use their power with reference to the evaluation can affect the power of evaluation. The constitutive power of evaluation is shaped and reinforced in all phases in terms of what is to be conceived as valid knowledge, reliable methodology and data.

Third, the framework reflects the power of the evaluation (system) in terms of its significance to support/impede key functions in public policy and governance. The power of evaluation depends on the relevance, quality, and timing of the evaluation and stakeholder involvement, factors known to enhance evaluation use (Cousins and Leithwood, 1986; Dahler-Larsen, 2015; Mark and Henry, 2004; Plottu and Plottu, 2009; Shulha and Cousins, 1997). How actors perceive these aspects of the evaluation affect whether and, if so, how they use their power with reference to the evaluation in governance. Actors can use their power to let the evaluation (system) support/impede the following functions: collective learning, democratic discourse, legitimizing policy/governance, policy improvement (e.g. extend or revise a policy or program), and democratic accountability. The power of evaluation can manifest as support for an evaluated policy or the legitimization of the applied governance model. For example, an evaluation of the effectiveness of public and private service organizations can reinforce or challenge new public management (NPM) governance (indicated by dotted upward arrow).

The constitutive and disciplinary power of evaluation (systems) can support the functions that the governance model relies on and at the same time provide little support to or impede other democratic governance functions.

Table 1 lists factors and conditions that can affect power relations and trigger actors’ use of power in the evaluation process. The governance structure, governance model, and evaluation policy affect power relations in all phases. How political directives and the evaluation contract are interpreted can trigger actors to use their power to act in the design phase. If the commissioner uses its power-over the evaluator and stakeholders to limit the evaluator’s

| Table 1. Conditions and factors affecting power in the evaluation process. |
|----------------------------------|----------------------------------|
| **All phases:**                  | - governance structure and governance model |
|                                  | - evaluation policy               |
| **Design phase:**                | - political directives            |
|                                  | - evaluation commission           |
|                                  | - design of evaluation/evaluation model |
| **Implementation phase:**        | - how the evaluation manager resolves emerging challenges |
|                                  | - how the evaluator manages (asymmetric) power relations and emerging challenges |
| **Conclusion phase:**            | - what commissioners conceive as valid knowledge, methodology, and data |
|                                  | - how the commissioner and stakeholders receive and respond to draft reports |
|                                  | - whether and how the evaluator takes into account comments, suggestions, and demands |
| **Communication phase:**         | - implementation of communication plan or decision on dissemination |
|                                  | - how the evaluation is reflected in briefings and other external communication |
|                                  | - media representations of the evaluation |
discretion or does not agree to modify evaluation questions, this can activate actors to mobilize their power-to act. How the commissioner and evaluation manager deal with emerging challenges can affect whether actors use their power-to act during the implementation phase. If the commissioner refuses to include data that reflect the negative consequences of a policy/program, an actor may decide to leave the evaluation. How the commissioner responds to the final draft report, in the conclusion phase, can affect whether/how the evaluator and stakeholders use their power-to act. Whether and, if so, how the evaluator takes into account comments, suggestions, and demands can prompt the commissioner to use its power-over the evaluator and stakeholders to use their power-to act. Whether and, if so, how the evaluator uses her or his power-to act if, for example, a summary or a covering letter to the final report misrepresents the evaluation or if the media cover the evaluation unfairly.

Table 2 lists conditions and factors that can increase or decrease the power of evaluation. How well-known factors increasing the use of evaluation are perceived by decision-makers and other stakeholders can increase or decrease the use, and thus the power, of evaluation. If an evaluation supports the government’s endeavors and reinforces the governance model, decision-makers can be expected to use it. When they use their power to substantiate arguments for continuing a policy with reference to the evaluation, the power of evaluation increases. The power of evaluation also increases when actors support the constitutive effects of an evaluation. The evaluation (system) can, for example, reinforce and justify administrative tools and routines if these are not problematized or questioned in the evaluation.

**Applying the framework in empirical research**

It is suggested that empirical studies, based on the framework, can be carried out in five steps: (1) Case description, (2) exploration of power in evaluation, (3) exploration of power of evaluation, (4) exploration of the interplay of power in and of evaluation, and (5) interpretation of implications for democratic governance.

The framework does not prescribe specific methods. Participant observation, stakeholder interviews and questionnaires (ex nunc or ex post) can be used for data collection. Analysis of the manifestations of power can use some kind of interpretative policy analysis (Fischer et al., 2007; Yanow, 2000) and focus on prevailing manifestations of power and how power dynamics evolve and how evaluation and democratic governance interplay.
Applying the framework to an evaluation

In this article, an evaluation of a Swedish national teacher-training program, the “Literacy Lift,” is used to demonstrate how the framework can be applied in empirical research in five steps. Steps 1–4 are developed below and Step 5 in the “Discussion” section. The data collection methods used is participant observation and collection of policy documents. The participant observation, undertaken by the author as project leader, include observations of power when actors take action, no action, interact with and react to one another, and other ways power manifest. These observations and collected policy documents are analyzed as an interpretative policy analysis (Yanow, 2000) and a qualitative directed content analysis (Hsieh and Shannon, 2005) of the research questions.

The case

The Swedish government initiated the Literacy Lift national teacher-training program in 2013 in response to the country’s decade-long failings in the Program for International Student Assessment (PISA; Ministry of Education, 2013). The government commissioned the National Agency for Education (NAE) to develop the program. It started in 2015 and ended after the 2019–2020 school year for schools, but continued for preschool teachers during the 2020–2021 school year. The program mainly consists of a learning platform with around 40 web-based literacy modules; groups of teachers led by a supervisor work with two modules for one school year according to a structured training model. As of 2020, Literacy Lift has reached around 25 percent of Swedish teachers in preschool, compulsory, and upper-secondary education. The program’s aim is to enhance teachers’ continuing professional development through what is called “collegial learning” in literacy, to improve teaching and, ultimately, students’ literacy and Sweden’s scoring on future PISA tests.

After procurement, the NAE commissioned a group of researchers to undertake a real-time evaluation of the program. The evaluation, designed as an objectives-oriented stakeholder evaluation, has presented 13 interim reports and a final report. In short, the evaluation shows that the program has helped enhance teachers’ knowledge of and insight into literacy and helped improve the teaching of literacy, but it has not achieved some of the program’s ambitious goals, such as improving all students’ literacy and improving Swedish students’ performance in PISA. Nor has it succeeded in institutionalizing structures for continuing collegial learning and teaching about literacy in all subjects.

The NAE repeatedly reported program progress to the government using interim evaluation reports together with their own follow-ups. The government extended the Literacy Lift program in 2016 and the evaluation was extended as well. The staff responsible for the program and the evaluation at the NAE were changed three times, whereas the evaluators stayed the same. The agreement to develop empirically thick interim reports and a brief analytical final report based on the interim reports, made at the beginning of the evaluation, was questioned by the new evaluation manager and program staff. The new demand was to explicitly describe all the data, methods, and statistical analyses in the final report. The evaluators presented a final draft report in April 2019 without objection from the reference group, but the NAE requested additional statistical analysis of the result interpretations. The evaluators first refused to do this. If they had not agreed to do this, the report would not be accepted for publication and the remaining money for the commission would not be paid. The additional statistical analysis did not change any major results or conclusions, and the final report was
approved in October 2019 but not approved for publication. The NAE presented its own report to the government together with the final evaluation report in June 2020, and at the same time, the evaluation report was approved for publication.

Stakeholders in the program and evaluation comprise the previous and current government, the NAE, school owners, school principals, head-teachers, and teachers/preschool teachers, reference group members, evaluators, Swedish Association of Local Authorities and Regions (SALAR), and Swedish Association of Independent Schools (SAIS).

**How power manifests and evolves in the evaluation process**

In the *design phase*, the *government* used its power-over the NAE to determine the objectives and direction of the program, expressed in two government directives (Ministry of Education, 2013, 2016) and to commission the NAE to develop and implement the program and organize the evaluation. The *NAE* used its delegated power to design and develop the program, formulate terms of reference and procurement terms and conditions for the evaluation, and select the evaluators. The NAE’s power-over the evaluators, codified in the contract, specifies the evaluators’ obligations, deadlines, conditions for report publication, and so on. The *evaluators* were delegated power-to act in line with the evaluation proposal after the contract was signed. The project leader of the evaluation used his power-to interpret the contract and, together with the evaluators, to further plan the evaluation and organize and chair the reference group meetings. The *reference group* members used their power to give comments on the evaluation design and to advocate for what they thought needed special attention, which differed somewhat. They also used their power to ask critical questions of the NAE about what they considered program shortcomings. The reference group empowered the evaluators and strengthened the evaluators’ arguments in relation to the NAE.

As noted, the design phase of the evaluation continued during program implementation. Adjustments were continuously made in the evaluation and, for example, new case studies were developed. The NAE and the evaluators discussed adjustments and extensions in which both parties had power as to whether or not to accept adjustments and extensions. NAE’s managers and staff changed during program implementation and agreements made with the evaluators at an early stage dissolved when new managers and staff used their power to demand changes in the design of the final report. The discussion about extending the evaluation to cover the extended Literacy Lift program reflects a brief moment of transformative power—that is, the NAE’s and the evaluators’ power increased when their commissions expanded, benefiting both parties. The case illustrates the dynamic nature of power, that is, how power manifests as power-over, power-to act, and transformative power at the beginning of the evaluation process.

During the *implementation phase*, the *evaluators* used their discretion and power-to act to implement the evaluation according to their proposal and contract and to develop feasible solutions to challenges related to, for example, data collection and case selection for in-depth studies. The evaluation model, that is, an objectives oriented effect and stakeholder evaluation, provided the *evaluands* with power to assess the performance and effects of the program. The evaluators used their power to analyze the evaluands’ assessments of program effects guided by evaluation, implementation, and literacy research.

The NAE used its power-over the evaluators to request recurrent feedback on the progress of the evaluation and to assess the quality of the draft reports. The NAE frequently
asked the evaluators to add coverage of action taken to improve the modules after the first data were collected. The evaluators used their power to judge the relevance of the NAE’s and the reference group members’ comments and suggestions and to decide what to modify, delete, or add in the draft reports. Not all evaluators always perceived the comments and how to address them in the same way. The project leader of the evaluation used his power-over colleagues to suggest how to manage the NAE’s comments, what changes to make in draft reports, and what analysis could be saved for the final report. Two stakeholders, that is, SALAR and SAIS, empowered themselves and claimed to have the right to comment on and demand changes in the electronic questionnaires. How these two organizations obtained this power-over the NAE and the evaluators is unknown. In any case, the evaluators had to obtain approval before the questionnaires could be distributed. The NAE’s power-over the evaluators during the implementation phase was mostly veiled but became obvious when the NAE repeated its comments on result interpretation and what should or should not be highlighted in the interim reports.

In the conclusion phase, the NAE used its power-over the evaluators to insist that the evaluators not only complement the final report as described earlier, but also revise some conclusions and recommendations. The NAE argued that conclusions and recommendations should take into account that the shortcomings pointed out in earlier reports had already been addressed. The evaluators used their power to resist pressure to adjust findings as requested by the NAE, claiming that the conclusions must be based on data that the evaluators had collected and that improvements made after data collection were considered when interpreting and discussing findings. How to deal with this is a recurrent challenge in real-time evaluation and a situation in which power can manifest. The reference group used its power to offer interpretations and explanations of findings, relate findings to different literacy research, and comment on weaknesses in the program. The evaluators used their power to decide what comments and suggestions to take into account when completing the reports.

In the communication phase, the NAE’s power-over the evaluators, as described, clearly manifests in the decision about disseminating the final report. The power-over the evaluators, codified in terms of reference, specifies that all dissemination and communication of the evaluation should be agreed to by the NAE before the evaluation is terminated. The Agency used its power-over the evaluators to withhold the decision on publication. The evaluators unsuccessfully used their power to have the report published, arguing that the value of the evaluation would be diminished for the stakeholders if the evaluation were not published when it was ready for publication.

**How power of evaluation manifests**

The power of the evaluation manifests in several ways. As indicated, the evaluation contributed to program development even during program implementation, showing that the power of real-time evaluation can manifest as support for the collective learning and policy improvement functions. The significance of the evaluation is also reflected in the NAE’s own report to the government. The Agency made comments on the evaluators’ recommendations, stating that it largely agreed with them (NAE, 2020). The government’s decision to continue the program for preschool teachers is in line with one of the evaluators’ recommendations, illustrating the power of evaluation for legitimizing program continuation. Another recommendation was to continue developing the modules and make them easily accessible, which the NAE
wrote that it would do, indicating the power of evaluation as providing ongoing support for the policy improvement function.

The training model, that is, the systematic work procedures for the modules, of which the government and NAE expressed high expectations, was questioned in the evaluation. The evaluators expressed doubts that the model could be the training model for Swedish continuing teacher training. The training model reflects how knowledge-steering and disciplinary power manifests in this case. When the evaluators questioned the model’s relevance as a blueprint for all teacher training, this diminished the disciplinary power of the model.

A constitutive effect of the evaluation is that it reinforced the NAE’s notion of what defines valid and reliable knowledge, that is, knowledge based on statistically significant results (a p-value of 0.05 is, by general convention, set as the cutoff for statistical significance). Although the evaluators argued that the evaluation was based on a pluralistic epistemology involving triangulation and that they considered data derived from interviews, questionnaires, and observations equally valid, the NAE demanded additional statistical analysis and the highlighting of statistically significant results in the final report. The NAE, supported by senior managers, used its power-over the evaluators—pluralistic scholars who considered different kinds of knowledge and evidence as valid for drawing conclusions on program performance and effects—to do something that they would not otherwise do, demonstrating how power-over can manifest and affect the evaluation report and the power of evaluation.

Relation between power in and of evaluation

The relation between power in and of evaluation manifested through the government’s and the NAE’s governance model and evaluation policy, which shaped the evaluation process and the functions the evaluation should support in national education policy and governance. A relation also manifested through the NAE’s and the reference group’s use of power to influence the evaluation and the final report and through NAE’s report to the government with reference to the evaluation. This confined the power of the evaluation to support the policy improvement and legitimatizing functions.

Conclusion

The framework contributes to enhance understanding and knowledge of power in and of evaluation and of the relation between the “in” and “of” in democratic governance.

The application of the framework shows how power manifests in the evaluation process and how the commissioner’s power-over the evaluator and the evaluator’s power-to act are intertwined, triggering each other. The NAE’s power-over the evaluator was mainly veiled but manifested openly when the Agency could not induce the evaluators to do what it wanted them to do. For a brief moment, power also manifested as transformative power.

The power of the evaluation manifested as support for the policy improvement and legitimatizing functions. The final evaluation report was used by the NAE as support to legitimate continuation of the program for preschool teachers. The power of the evaluation diminished when the NAE used its power-over the evaluators to postpone publication of the final evaluation report, which blocked the evaluators’ power-to act and held back the evaluation’s ability to support collective learning and democratic discourse.
The case also illustrates the constitutive power of the evaluation in shaping the notion of what is valid and reliable knowledge. The NAE’s demands helped reinforce and constitute the notion of statistically significant knowledge as the most valid type of knowledge. The case also illuminates the program’s disciplinary power and how the evaluation reduced this power when questioning the feasibility of the training model to serve as a blueprint for continuing teacher training. Hence, the case demonstrates how the NAE’s and the evaluators’ mobilization of power affected the constitutive and disciplinary power of the evaluation and the program.

The relation between power in and of evaluation manifested through the government’s and the NAE’s governance model and evaluation policy, which shaped the evaluation process and paved the way for the functions that the evaluation was intended to support in national education policy and governance. The relation also manifested through the NAE’s and the reference group’s use of power to comment on the evaluation design and draft reports and in the NAE’s reports to the government when interpreting the evaluation and in holding back publication. This affected the final report, its usability and eventually diminished the power of the evaluation to support further democratic governance functions than the policy improvement and legitimizing functions.

Discussion

The framework takes into account the four shifts away from research-based ad hoc evaluation informing decision-makers in representative democracy (Dahler-Larsen, 2021) and other trends that have changed the conditions for evaluation (Hanberger, 2018; Picciotto, 2015). To what extent institutionalized evaluation systems replace stand-alone evaluations and decision-makers lose democratic control over evaluation are empirical questions. The case demonstrates how power in evaluation manifests as supporting political and administrative decision-making and reflects more rather than less administrative and democratic control. The NAE continues to commission stand-alone evaluations at the same time as the government, the NAE, quasi-governmental and private actors have institutionalized at least 30 evaluation systems for primary and secondary education in Sweden (Benerdal, 2019; Lindgren et al., 2016). Thus, stand-alone evaluations and various overlapping monitoring and evaluation systems operate in a crowded policy space in the Swedish education system. Decision-makers tend to lose control over evaluation systems (Andersen, 2021; Dahler-Larsen, 2021; Lindgren et al., 2016) but as this case shows not over stand-alone evaluation. Further research can contribute to knowledge of whether stand-alone evaluations are replaced by evaluation systems in different policy fields and governance models and if, and if so how and why, administrative and democratic control is affected.

The case demonstrates the multiple and subtle ways power can manifest and evolve in the evaluation process. For example, new actors can enter the scene and empower themselves and act as power-holders, as two stakeholders (i.e. SALAR and SAIS) did over the NAE and the evaluators. These actors’ empowerment illustrates how NPM governance affect evaluation (Hanberger, 2012; Schoenefeld and Jordan, 2017), shaping new “political opportunity structures” (Schoenefeld and Jordan, 2019) that provides quasi-governmental actors power to influence the evaluation. The case also demonstrates how the NAE shaped the constitutive power of the evaluation when it used its power-over the evaluators to reinforce the conception of statistical significance-based knowledge as the most valid type of knowledge (Andersen, 2020; Dahler-Larsen, 2012). This could be a strategy to avoid a polarized and mediatized
For an evaluation to effectively change another practice, in this case, to contribute to improve teaching in literacy and teacher training (referred to as the policy improvement function in the framework), it must protect itself from contestability and be backed up by trust in the methodology, the data, and the institution that carries out evaluation (Dahler-Larsen, 2015). The NAE used its power-over the evaluators not only to ensure trust in the methodology of the evaluation but also to build trust in the NAE as a competent knowledge-steering agency. With reference to the evaluation, it developed the program and governed local school actors to implement the program according to the training model. The power of the evaluation to change teaching practice was intertwined with the power of the program. However, most important to change teaching practice during the program was the application of the training model, supported by a state subsidy, the NAE’s demands on participants, and its control of program implementation. The NAE used the evaluation to refine the program and improve program implementation indicating some power of the evaluation to support developing teaching practice. The power of evaluation can also manifest after the program. However it is not known if the program material and the evaluation reports, which are free to use and accessible from the NAE’s website, are used by teachers and head-teachers in developing teaching practice after the program.

The framework recognizes that public-sector evaluation is confined by and affects governance (Step 5). The facts that the previous and current government have power to develop and change the conditions for evaluation and the NAE has power to develop and revise terms of reference for evaluation reflect the division of power between the political and administrative levels in representative democracy and the state-model of governance (Hanberger, 2009). The NAE used its delegated power to manage the evaluation to match the policy cycle, which diminished the power of the evaluation to support further functions in democratic governance. The NAE used the evaluation to reinforce elite-democratic governance at the same time as it blocked the collective learning and discursive democratic functions (Habermas, 1996; Hanberger, 2006). The evaluation was subjected to the NAE’s evaluation policy (Habermas, 1996; Hanberger, 2006; Trochim, 2009) and response system (Hanberger, 2011) thus confined by administrative power (Widmer and Neuenschwander, 2004). The project leader for the evaluation used his power-to act in the evaluation process and managed some but not all power imbalances (Baur et al., 2010). He failed to get the report published when finalized which blocked the evaluators’ opportunity to support the collective learning and democratic discourse functions.

The framework treats power in and of evaluation as related issues embedded in the same governance structure and governance model. Andersens’ (2021) observation that an evaluation system both can increase power-over evaluands and decrease the power of evaluation reflects a relation between power in and of evaluation. The relation between the “in” and “of” also depends on whether actors in the evaluation process use their power to influence the evaluation, if the evaluator let comments and suggestion affect the evaluation (system) and if any of these actors refer to the evaluation to support functions in democratic governance. However, if an evaluation (system) is not referred to at all in the continuing policy process or policy discourse, there is no actor-based relation between the “in” and “of,” but if participants in the evaluation process refer to the evaluation later on, it reflects a delayed actor-based relation between the “in” and “of.”
The framework’s advantage is that it contributes to a comprehensive, multifaceted, and dynamic understanding and exploration of power in and of evaluation in democratic governance and of the relation between the “in” and the “of.” It can be used conceptually or in empirical research as a whole or solely to focus on the “in” or “of.” The exploration of power, in this case, is based on the author’s observations and interpretations. If more actors’ observations and experiences are collected, the representation of power becomes more multifaceted.

The framework does not support the exploration of power in different evaluation models/approaches specifically and not of the power of evaluation in general. A limitation is also that the framework does not pay attention to other contextual factors (Haugen and Chouinard, 2019). For example, the community context, the evaluator’s and other actors’ competence, experience, and knowledge factors that can affect the evaluation products and the power of evaluation.

Another limitation, when using the framework empirically, is that power is by nature difficult to explore; actors may not want to share why they use or do not use their power because this would reveal information that is meant to be kept secret.

As indicated how power manifests differ not only between but also within stakeholder groups/organizations. NAE’s senior managers, evaluation managers and literacy staff did not use their power in the same way in interaction with the evaluators. Members of the reference group also used their power somewhat differently to comment on the program and evaluation. How power manifests in shaping and utilizing political opportunities in evaluation and in promoting (organizational/collective) learning (Schoenefeld and Jordan, 2019) is intertwined, multifaceted, and further research should reflect power differentials both between and within stakeholder groups/organizations.

To enhance understanding and knowledge of what affects the power of evaluation the factors and conditions listed in Table 2 and other functions that could increase or decrease the power of evaluation to support key governance functions merit further research in different policy fields and governance contexts.

If Andersen’s (2021) recognized paradox of evaluation systems also appears in, for example, evaluation systems that combine indicator-based monitoring systems and recurrent in-depth evaluations, and how constitutive and disciplinary power affects the evaluation process and the power of evaluation, and how actors can affect these kinds of power merit further research in different policies and governance contexts.

It is hoped that this article can inspire further research and discussion of power and evaluation while the conditions of evaluation and democratic governance are changing.

Acknowledgement
I thank the anonymous reviewers for their careful reading of my manuscript and their many insightful comments and suggestions.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.
Notes
1. Evaluation is used broadly referring to both stand-alone evaluation and evaluation systems.
2. Governance refers to new emerging institutions that have been transformed from established forms of government or that have evolved independently of the state, including different modes of governing and coordination where public and private institutions and actors are involved (Björk et al., 2003; Pierre and Peters, 2000).

References
Allen A (1999) The Power of Feminist Theory: Domination, Resistance, Solidarity. Boulder, CO: Westview Press.
Allen A (2016) Feminist perspectives on power. In: Zalta EN (ed.) The Stanford Encyclopedia of Philosophy (Fall 2016 Edition). Available at: https://plato.stanford.edu/archives/fall2016/entries/feminist-power/ (accessed 10 September 2020).
Andersen N (2020) The constitutive effects of evaluation systems: Lessons from the policymaking process of Danish Active Labour Market Policies. Evaluation 26(3): 257–74.
Andersen N (2021) The power and paradoxes of evaluation systems—Increasing use but impeding change. Scandinavian Journal of Public Administration 25(3/4): 39–59.
Arendt H (1970) On Violence. New York: Harcourt Brace & Co.
Bachrach P and Baratz M (1962) Two faces of power. American Political Science Review 56: 947–52.
Ball T (1992) New faces of power. In: Wartenberg T (Ed.) Rethinking Power. Albany, NY: State University of New York Press, 14–31.
Barnes B (1988) The Nature of Power. Oxford: Polity.
Baur V, Van Elteren A, Nierse C, et al. (2010) Dealing with distrust and power dynamics: Asymmetric relations among stakeholders in responsive evaluation. Evaluation 16(3): 233–48.
Benerdal M (2019) Nationell Utvärderingspolicy: Utformning Och Förändring På Grundskoleområdet 1988–2014 [National Evaluation Policy: Its design and change in compulsory education from 1988–2014]. Doctoral Dissertation, Umeå, Umeå University.
Björk P, Bostedt G and Johansson H (2003) Governance. Lund: Studentlitteratur.
Bovens M, ’T Hart P and Kuipers S (2006) The politics of policy evaluation. In: Moran M, Rein M and Goodin RE (eds) The Oxford Handbook of Public Policy. Oxford: Oxford University Press, 319–35.
Chelimsky E (2009) Integrating evaluation units into the political environment of government: The role of evaluation policy. New Directions for Evaluation 123: 51–66.
Cousins B and Leithwood K (1986) Current empirical research on evaluation utilization. Review of Educational Research 56(3): 331–64.
Dahl R (1957) The concept of power. Behavioral Science 2: 201–15.
Dahler-Larsen P (2012) The Evaluation Society. Stanford, CA: Stanford Business Books.
Dahler-Larsen P (2014) Constitutive effects of performance indicators: Getting beyond unintended consequences. Public Management Review 16(7): 969–86.
Dahler-Larsen P (2015) The evaluation society: Critique, contestability, and skepticism. Spazio Filosofico 1(13): 2136.
Dahler-Larsen P (2021) The circulation of evaluative information. Scandinavian Journal of Public Administration 25(3/4): 17–35.
Digeser P (1992) The fourth face of power. The Journal of Politics 54(4): 978–1007.
Eckhard S and Jankauskas V (2019) The politics of evaluation in international organizations. A comparative study of stakeholder influence potential. *Evaluation* 25(1): 62–79.

Fischer F, Miller G and Sidney M (2007) *Handbook of Public Policy Analysis: Theory, Politics and Methods*. New York: Routledge.

Foucault M (1977) *Discipline and Punish: The Birth of the Prison* (trans. A Sheridan). New York: Vintage.

Furubo JE and Karlsson Vestman O (2011) Evaluation for public good or professional power. In: Eliadis P, Furubo J-E and Jacobs S (eds) *Evaluation: Seeking Truth or Power?* New Brunswick, NJ: Transaction: Publishers, 1–35.

Habermas J (1996) Three normative models of democracy. In: Benhabib S (ed.) *Democracy and Difference: Contesting the Boundaries of the Political*. Princeton, NJ: Princeton University Press, 21–30.

Hanberger A (2006) Evaluation of and for democracy. *Evaluation* 12(1): 17–37.

Hanberger A (2009) Democratic accountability in decentralized governance. *Scandinavian Political Studies* 32(1): 1–22.

Hanberger A (2011) The real functions of evaluation and response systems. *Evaluation* 17(4): 327–49.

Hanberger A (2012) Framework for exploring the interplay of governance and evaluation. *Scandinavian Journal of Public Administration* 16(3): 9–28.

Hanberger A (2018) Rethinking democratic evaluation for a polarised and mediatised society. *Evaluation* 24(4): 382–99.

Haugaard M (2010) Power: A “family resemblance” concept. *European Journal of Cultural Studies* 13(4): 419–38.

Haugen J and Chouinard JA (2019) Transparent, translucent, opaque: Exploring the dimensions of power in culturally responsive evaluation contexts. *American Journal of Evaluation* 40(3): 376–94.

Howlett M (2014) Why are policy innovations rare and so often negative? Blame avoidance and problem denial in climate change policy-making. *Global Environmental Change* 29: 395–403.

Hsieh H and Shannon S (2005) Three approaches to qualitative content analysis. *Qualitative Health Research* 15(9): 1277–88.

Klijn E (2008) Governance and governance networks in Europe: An assessment of 10 years of research on the theme. *Public Management Review* 10(4): 505–25.

Kothari U (2001) Power, knowledge and social control in participatory development. In: Cook B and Kothari U (eds) *Participation: The New Tyranny?* London: Zed Books, 139–52.

Leeuw F and Furubo JE (2008) Evaluation systems: What are they and why study them? *Evaluation* 14(2): 157–69.

Lindgren L, Hanberger A and Lundström U (2016) Evaluation systems in a crowded policy space: Implications for local school governance. *Education Inquiry* 7(3): 237–58.

Lukes S (2005) *Power: A Radical View*, 2nd expanded edn. London: Macmillan.

Mark M and Henry G (2004) The mechanisms and outcomes of evaluation influence. *Evaluation* 10(1): 35–57.

Ministry of Education (2013) (Government directive) *Uppdrag om fortbildning i läs—och skrivutveckling—Läslyftet* (Regeringsbeslut U2013/7215/S, 2013-12-05). Stockholm: Utbildningsdepartementet.

Ministry of Education (2016) (Government directive) *Tilläggsuppdrag om fortbildning i språk—, läs—och skrivutveckling—Läslyftet* (Regeringsbeslut U2016/05733/S, 2016-12-22). Stockholm: Utbildningsdepartementet.

National Agency for Education (NAE) (2020) *Redovisning Av Uppdrag Om Fortbildning I Läs—Och Skrivutveckling—Läslyftet*. Stockholm: Skolverket.

Picciotto R (2015) Democratic evaluation for the 21st century. *Evaluation* 21(2): 150–66.

Pierre J and Peters G (2000) *Governance, Politics and the State*. New York; Basingstoke: Macmillan.
Plottu B and Plottu E (2009) Approaches to participation in evaluation: Some conditions for implementation. *Evaluation* 15(3): 343–59.

Raimondo E (2018) The power and dysfunctions of evaluation systems in international organizations. *Evaluation* 24(1): 26–41.

Rutkowski D and Sparks J (2014) The new scalar politics of evaluation: An emerging governance role for evaluation. *Evaluation* 20(4): 492–508.

Rye D (2015) Political parties and power: A new framework for analysis. *Political Studies* 63(5): 1052–1069.

Schoenefeld J and Jordan A (2017) Governing policy evaluation? Towards a new typology. *Evaluation* 23(3): 274–93.

Schoenefeld J and Jordan A (2019) Environmental policy evaluation in the EU: Between learning, accountability and political opportunities? *Environmental Politics* 28(2): 365–84.

Shulha L and Cousins B (1997) Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice* 18(3): 195–208.

Trochim WM (2009) Evaluation policy and evaluation practice. *New Directions for Evaluation* 123: 13–32.

Wartenberg T (1990) *The Forms of Power: From Domination to Transformation*. Philadelphia, PA: Temple University Press.

Weber M (1978) *Economy and Society: An Outline of Interpretive Sociology* (Trans. E Fischoff, et al.) Berkeley, CA: University of California Press.

Weiss C (1993) Where politics and evaluation research meet. *Evaluation Practice* 14(1): 93–106.

Weiss C (1999) The Interface between evaluation and public policy. *Evaluation* 5(4): 468–86.

Widmer T and Neuenschwander P (2004) Embedding evaluation in the Swiss federal administration: Purpose, institutional design and utilization. *Evaluation* 10: 388–409.

Yanow D (2000) *Conducting Interpretive Policy Analysis*. Newbury Park, CA: SAGE.

Anders Hanberger is Professor Emeritus at Umea University, Sweden. He has published articles on democratic evaluation, evaluation of public policies and programs, evaluation systems, evaluation methodology, and policy analysis.