

Anders Hanberger
Umea Centre for Evaluation
Research, Umea University

Keywords: evaluation
steering, evaluation use,
governance, functions of
evaluation, framework,
program theory

Nyckelord: utvärderingsstyr-
ning, governance, utvärde-
ringsanvändning, utvärde-
ringsfunktioner, analysram,
programteori

Abstract

The article develops a conceptual framework for the interplay between evaluation and governance and explores key functions of evaluation in democratic governance, mainly the accountability and improvement function. Using this framework contributes to knowledge of how governance affects evaluation and evaluation systems, and how evaluation may contribute to policy or governance change. Functions of evaluation are explored as effects of evaluative information produced by different evaluations on the one hand, and as evaluation system effects on the other. The framework can be used to map out and explore the interplay between evaluation and governance in different policy sectors. Program theory methodology is used to unpack the assumptions underlying evaluation in the state and network models of governance. The program theories can be used to illuminate and discuss how evaluation should work in different models of democratic governance, as well as for empirical research of the interplay between governance and evaluation.

Att studera samspelet mellan utvärdering och governance

I artikeln utvecklas en analysram för att studera samspelet mellan utvärdering och governance och centrala funktioner som utvärdering kan fylla i demokratisk governance, främst ansvarsutkrävande och förbättring. Begreppen governance och demokratisk governance refererar till olika former av samhällsstyrning. Analysramen bidrar till att utveckla kunskap om hur governance påverkar utvärdering och utvärderingssystem och hur utvärdering kan förändra policy och governance. Utvärderingars funktioner studeras i artikeln dels som effekter av kunskap producerad i olika utvärderingar, dels som effekter av utvärderingssystem. Analysramen kan bland annat användas för att utforska samspelet mellan utvärdering och governance inom olika policysektorer. Antaganden som finns bakom statlig governance respektive nätverks governance analyseras med hjälp av programteorimetodologi. De framtagna programteorierna kan användas för att synliggöra och diskutera hur utvärdering bör fungera i olika former av governance och för att studera samspelet mellan utvärdering och governance i praktiken.

Offentlig förvaltning
Scandinavian Journal of
Public Administration
16(3): 9-27

© Anders Hanberger and
Förvaltningshögskolan, 2013
ISSN: 1402-8700
e-ISSN: 2001-3310

***Anders Hanberger** is Professor in Evaluation and director of research at Umea Centre for Evaluation Research, Umea University. He has conducted evaluations for government agencies and other organisations, and is engaged in research on evaluation and evaluation systems.

Introduction

Monitoring and evaluation (M&E) are highly regarded in today's society, which has led to increasing demand for them and to great expectations regarding what they will deliver. It is assumed that M&E support democratic governance and promote accountability and policy and program improvement (Weiss, 1999; Schwandt, 2002; Mark & Henry, 2004; Hansson, 2006; Pollitt, 2006). However, research has shown that when it comes to the governance of public policy, M&E also have other functions, some of which are symbolic (Boswell, 2009; Hanberger, 2011; Power, 1997, 2007; Schneider & Ingram, 1997). M&E may also have unexpected effects and consequences (Alderson & Wall, 1993; Cousins & Leithwood, 1986; Cousins et al., 2004; Hanberger, 2011; Hansson, 2006; Lindgren, 2006, 2011; Leeuw & Furubo, 2008; Perrin, 1998; Plottu & Plottu, 2009; Schaumburg-Müller, 2005; Shulha & Cousins, 1997; Strathern, 2000; Valovirta, 2002; Weiss, 1979). There is clearly a dearth of knowledge regarding how M&E work and function in different models of governance. This article contributes to knowledge of the interplay between evaluation and governance, and of the functions of evaluation systems in democratic governance.

With evaluators, inspectors, NGOs, the media, and other reviewers monitoring and evaluating those in power, current programs, and policies, we have arrived at a situation of overlapping evaluations and an abundance of evaluative information. Policy makers interpret the value and implications of different evaluations before focusing on a single evaluation and using it to improve policy. A problem is that "in the real world, multiple evaluations of the same policy tend to be non-cumulative and non-complementary. Their methods and findings diverge widely, making it hard to reach a single authoritative or at least consensual judgement about the past and draw clear-cut lessons from it" (Bovens, 't Hart & Kuipers, 2006:321).

Due to this situation and to the political implications of evaluations, those concerned with and affected by an evaluation not only pay attention to its actual findings but also to other factors such as the questions that were not raised, ulterior motives, validity issues, and the negative effects of measurements. In the broader policy and accountability environment, the findings are discussed together with other knowledge and endeavours. The political, social, and cultural implications of evaluation are interpreted in a context in which the wider implications and functions of evaluation are considered equally important (Behn, 2001; Bovens, 't Hart & Kuipers, 2006; Boswell, 2009; Klijn, 2008; Dahler-Larsen, 2007, 2010; Hanberger, 2011; Vedung, 1997).

If more attention is paid to the governance structure in which evaluation is embedded, we can arrive at a better understanding of the implications of evaluation in public policy and governance (Bovens, 't Hart & Kuipers, 2006; Cousins & Leithwood, 1986; Hertting & Vedung, 2012). There is also a need for discussion of how to study the implications of evaluation in the context of governance, and for development of approaches to studying the implications of evaluation and the interplay of evaluation and governance.¹ Generally, M&E might contrib-

ute to improving existing programs and policies and maintaining the established governance model, or they can provide new insights to be considered in the policy and governance discourse that can eventually contribute to change in policy and governance. However, how evaluation works in the real world of governance is still a scarcely explored domain.

The purpose of this article is to discuss the interplay of evaluation² and governance and to develop a conceptual framework for exploring this interplay and key functions of evaluation in democratic governance. The focus is on two functions of evaluation, namely accountability and improvement, but the framework also takes into account four further functions that evaluation may have in democratic governance, namely, critical, learning, legitimization, and symbolic functions. As the functions of evaluation are not clearly articulated in democracy and governance theories, there is a need to suggest how evaluation can be assumed to work in different models of governance. Therefore this article also explores assumptions in relation to two models of democratic governance (the state/ local/regional model and the network model), and discusses how evaluation steering can support the two models.

The article begins with a discussion of evaluation steering in relation to different models of governance. Next, key functions of evaluation in democratic governance are discussed in relation to three models of decentralized governance. A framework is then developed based on the previous discussion and theories, and two program theories for evaluation in democratic governance are developed. Finally, the article discusses implications of the framework for democratic governance and evaluation practice.

Theoretical considerations

Evaluation steering

Evaluation steering,³ that is, how evaluation is set up to meet the needs of governance, is based on different assumptions and involves different steering mechanisms in different models of governance. In new public management (NPM), for example, evaluation steering implies that M&E meets the needs of NPM, or in other words, promotes transparency, performativity, and accountability (Hansen, 2008; Pollitt, 1995; Power, 1997; Rist, 2006; Rhodes, 1997; Vedung, 2010). Adapting evaluation to serve democratic governance may mean different things, for example, steering evaluation to serve the needs of the political and administrative elite or the needs of the wider policy community (Chelimsky, 2006; Hanberger, 2006; House & Howe, 1999; MacDonald, 1976, 1978; MacTaggart, 1991; Ryan, 2004; Keane, 2008).

In the present special issue, Hanne Foss Hansen uses the term “systemic evaluation governance” to identify evaluation steering in a broader field:

SEG [Systemic evaluation governance] is defined as evaluation carried out with steering ambitions and targeted at several actors in a field. SEG for example may be targeted at several organisations in an organisational field, e.g., a policy sector, as is the case when educa-

tional institutions are benchmarked aiming at supporting free user choice, or it may be targeted at several nation states, as is the case in relation to the Open Coordination Method at the European Union level (Foss Hansen, 2012: 48).

Evaluation steering can also be explored in relation to the governance doctrines behind four different waves of evaluation (Vedung, 2010). For example, in the neo-liberal wave (NPM) evaluation has become a “permanent feature of results-based management and of outsourcing” (ibid., 273), and in the evidence-based wave, evaluations, meta-analysis and systematic reviews (p.274) are used to compile evidence-based knowledge to promote evidence-based practice (cf. Hansen & Rieper, 2009). Evaluation steering has facilitated the evolution of evaluation waves and contributed to maintaining governance doctrines. As indicated, evaluation steering implies different things in different models of governance.

Functions of evaluation in democratic governance

A key function of evaluation in governance is the promotion of democratic accountability. Democratic accountability is many-sided, for there are many principal-agent relationships in a representative democracy (Behn, 2001; Hanberger, 2009; Mulgan, 2003). Hanberger (2009)⁴ discusses democratic accountability in relation to three models of decentralized governance, and shows how evaluation can be used to support democratic accountability. He highlights how the governance structure frames and affects evaluation.

In the state model, elements of a state’s strong administrative system are used to implement state policies and programs through municipalities (cf. Lindvall & Rothstein, 2006:50). State inspectors, auditors, and evaluators evaluate (review) the implementation and performance of state policy on behalf of elected representatives. These evaluations can be used to hold lower levels of government and administration to account. Democratic control can also take place when citizens vote in general elections for a government to stay in power or vote for candidates from opposition parties. Hence, when holding governments to account, elected representatives and citizens can use all kinds of evaluative information regarding how the government has performed its tasks, implemented policies and programs, exercised its power and so forth. This model is in play when elected representatives constitute the government, when the government holds lower levels of government to account, and when citizens also act as accountability holders when voting.

The local-regional model relates to decentralized governance emerging from local and regional government. Discretion with regard to local and regional policy making can be more or less limited by the state, and may vary between policy sectors (cf. Lidström, 2004; Montin, 2002). Municipal auditors⁵ monitor local government and municipal policy on behalf of the elected representatives in the Swedish system and thus act as agents for the local or regional assembly. They can also act directly as reviewers for the citizenry. Citizens can control local and

regional government directly in elections, and citizens' organizations can also undertake reviews to hold those in power to account for local or regional policy. Democratic accountability is at work when elected officials use municipal audits to hold local governments to account for promises made, and when citizens use these audits to exercise their own democratic control.

In the multi-actor or network model, the national and local government share power with other actors, and join networks and partnerships in order to resolve pressing problems and challenges. This model, sometimes referred to as network governance (Hertting and Vedung, 2012), also includes multi-level governance in which power is shared between levels of government and between non-governmental actors or institutions, and thus also includes governance in the civil and private sphere. When assessed against the prerequisites for accountability in representative democracy, the conditions for accountability are weak in this model (cf. Weber, 1999: 482). But if the conditions for accountability are interpreted from a discursive democratic perspective, this model is the only one that addresses new forms of democratic accountability (Behn, 2001). An evaluation that supports democratic accountability in network governance takes into account key actors' knowledge and information needs. It should thus use a variety of evaluation criteria, including different actors' criteria of fairness and success. The function of democratic accountability is in play when stakeholders and citizens use evaluation as a source of information in holding those responsible to account, including politicians, administrators, and professionals.

The above paragraphs indicate that the evaluative information that facilitates accountability differs in the three models of decentralized governance.

Policy and program improvement⁶ is another key function of evaluation in democratic governance. The three governance models can also be used to illustrate how the policy improvement function can be promoted by evaluation.

The state model needs evaluative information concerning how policy instruments work. These instruments include regulative instruments (e.g., laws and rules and regulations), economic instruments (e.g., government grants), and information instruments (e.g., program objectives). Likewise, the local-regional model needs evaluative information about policy instruments developed by local governments. Evaluations that promote policy improvement in multi-actor or network governance should focus on such things as how collaboration and collective action works (Hertting & Vedung, 2012).

Evaluation can also have a critical function, that is, it can promote deliberation regarding a policy's core assumptions, objectives, or instruments, or about the governance model adopted. This function can be promoted by independent evaluation researchers (Määttä & Rantala, 2007).

Evaluation can also promote learning in governance. The learning function is distinctive and implies that the evaluation process is more or less integrated into the policy process. In other words, learning takes place continuously in a learning organization (Schaumburg-Müller, 2005). The state and local-regional model need evaluations to support elite learning, whereas the network model needs evaluations to facilitate collective learning.

Evaluation can also fill a legitimization function (Boswell, 2009; Hanberger, 2011), for example to help legitimize the governance model adopted or a policy that is being implemented.

If an evaluation (system) is treated as an artifact and used in a tokenistic way, it fills a symbolic function in governance.

Although evaluation can fulfill all these functions in democratic governance, the focus of this article is on the accountability and improvement function. The prerequisites for different functions differ, and a few are the same for almost all functions (e.g., evaluation should be of acceptable quality, accurate, relevant and credible). Hanberger's (2011) analysis of the common and specific prerequisites for various functions of evaluation (and response systems) in multi-actor policy making also informs the framework.

Framework

The framework developed on the basis of the above theories implies that the governance structure in which public policy and evaluation are embedded affects evaluation (Figure 1). It also recognizes that governance models may change as actors and institutions change, and that evaluation may contribute to policy and governance change.

A distinction is made between three types of evaluation that differ in structure, scope, and the knowledge produced. Type I evaluation refers to an M&E system that is indicator-based, producing continuous evaluative information on inputs, processes and outputs/outcomes of policies and programs. Type I evaluation may focus on the "M", in which case it mainly collects data on inputs and/or outputs for the purpose of monitoring implementation, or on the "E", in which case it focuses more on compiling outcome and performance measures indicating such things as achievement of objectives (for a further discussion, see Rist, 2006). In both cases the M&E systems produce streams of quantitative evaluative information (Stame, 2006).

Type II evaluation refers to stand-alone studies that generate quantitative and/or qualitative knowledge. A specific evaluation can make use of information produced by a type I evaluation, but consists of systematic inquiries, analyses, and assessments.

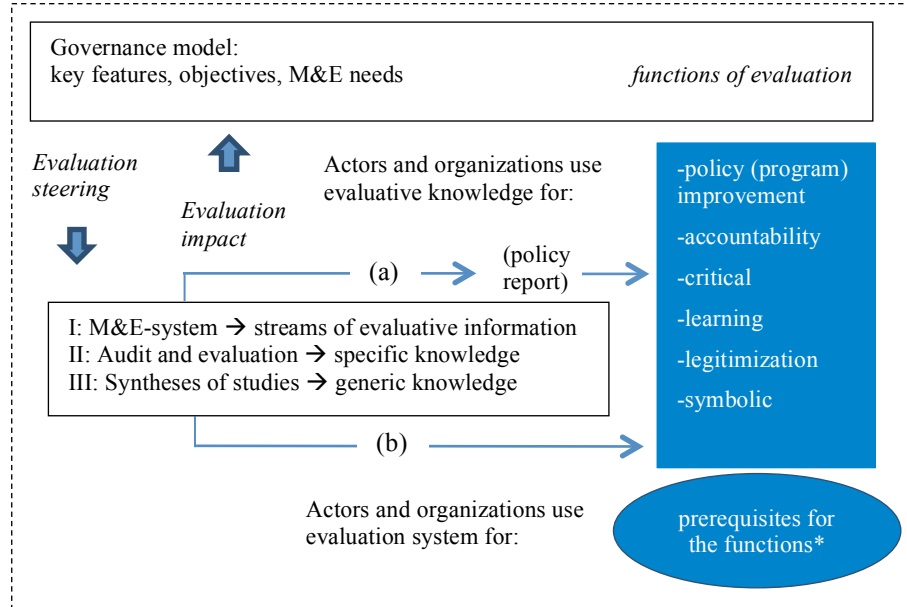
Type III evaluation refers to studies that synthesize and compile knowledge from a number of type II and I evaluations and other relevant studies.

The above categorization of evaluations differs somewhat from the four streams of evaluative knowledge discussed by other researchers (Stame, 2006: x). However, it enables us to explore the interplay of governance and evaluation, and the functions of evaluations in governance.

An evaluation system can be set up to maintain a governance structure or governance model, and within this structure or model it may fulfill an accountability and policy improvement function, for example. Thus the framework illustrated in Figure 1 takes into account that evaluation is framed by a governance structure (model) that confines and governs it. Evaluation can also have the main

function of maintaining or reforming the governance structure, that is, providing knowledge that eventually contributes to changing the governance structure (indicated by the upward pointing arrow). The governance structure to be explored may be a combination of the governance models discussed in the previous section or something different, such as an NPM model.

Figure 1: Framework to study functions of evaluation in democratic governance



*The specific prerequisites for fulfilling a particular function differ. A few prerequisites are the same for almost all functions. Most functions require that the evaluation should maintain acceptable quality, be accurate, relevant and credible (see Hanberger, 2011 for a discussion of the common and specific prerequisites for various functions).

When any evaluation or evaluation system is initiated, those in charge have expectations regarding what knowledge requirements it should meet and what functions it should serve. The intended functions may change over time or when governments change. Hence, an evaluation system may not always be guided by the original set of intentions, and may come to have different intended functions.

Figure 1 provides a conceptual framework for exploring the interplay of governance and evaluation. It presents key concepts and their relations, and allows for different analyses:

- Describing the features and objectives of a governance model and what knowledge it needs clarifies how the governance model embeds evaluation and the M&E needs.
- Unpacking program theories for evaluation in different models of governance contributes to normative and empirical knowledge of evaluation in governance.

- Exploring the real functions of evaluation in democratic governance contributes to knowledge of the implications of evaluation in governance.

The framework allows for two ways of exploring the functions of evaluation in governance: through the evaluative knowledge produced and through the system. Type I evaluation (M&E systems) produces ongoing evaluative information; type II (single studies) produces specific knowledge of a policy, program, or organization; and type III evaluation (synthesis studies) produces generic knowledge. The first way (arrow a) can be filtered through a policy report. The evaluation system can have functions such as legitimizing a certain governance model or justifying more state inspections in a policy domain (arrow b). The same evaluation or evaluation system can fulfill different functions for different actors and organizations.

The framework pays attention to six functions that evaluation can have in democratic governance (improvement, accountability, critical, learning, legitimization, and symbolic). The symbolic function is included to remind us that evaluation (systems) may eventually fill a symbolic or tactical function, but this function is not explored in this article (Hanberger, 2011). The functions that evaluation eventually has in the real world of governance depend on factors such as the governance structure, the effectiveness of evaluation steering, how actors and organizations interpret and use evaluation in relation to their own preferences and strategies, stakeholders' and organizations' engagement during the evaluation process, their trust in the knowledge produced, and anticipation of future evaluations.

Program theory for evaluation

It is important to establish the assumptions underpinning evaluation in different models of democratic governance because they are not the same in all models of governance. A public agency or organization may have an evaluation policy that makes the assumptions visible or it may be more or less veiled (cf. Hanberger, 2003, 2011).

One way of unpacking the assumptions underpinning evaluation is to apply a program theory methodology, that is, to reconstruct program theories for evaluation in democratic governance. The program theory consists of assumptions of how an evaluation system can achieve its intended effects in different models of governance.

It is recognized that different notions of democratic governance are based on different ideas about how accountability and policy improvement can be realized. Furthermore, governments, opposition parties, and stakeholders may not all share the same assumptions regarding how evaluation should contribute to policy improvement. Hence, there is a need to decide whose intentions and assumptions one should reconstruct (Leeuw, 2003).

Frans Leeuw (2006) illustrates how program theory methodology can be used to illuminate underlying assumptions regarding how evaluation systems can achieve their intended effects using the examples of the Dutch National Audit

Office (NAO), the Dutch Inspectorate for Education, and the Research, Statistics and Information Centre of the Ministry of Justice. The first two are briefly discussed here.

A program theory was reconstructed based on a significant number of studies of the type of evaluative knowledge the NAO collects for auditing. The theory makes explicit the NAO's assumptions about how a governmental actor can operate to achieve effective policy through responding and taking action in response to audits. Briefly, the NAO assumes that in order to be an effective and efficient policy actor, a public agency should comply with rules and regulations while developing and implementing a policy or program, should continuously act as a rational and learning actor in order to optimize implementation, and should improve conduct and routines if criticized by auditors (p. 86). This program theory focuses on organizational mechanisms, including information systems, coordination, cost-benefit analysis, planning, documentary evidence and audits (p. 85–87). Leeuw emphasizes that the Dutch NAO does not base auditing on research, which brings to the fore the importance of leadership in public organizations, the attitudes and cultures within them, and trust and commitment. The latter factors are at least “as important as the administrative-procedural factors that form the focus of the National Audit Office” (ibid., 87). The NAO's program theory for auditing thus has limited theoretical support.

The Dutch Inspectorate for Education evaluates the quality of the Dutch education system. Leeuw focuses on assumptions regarding the way its inspections operate at the school level. Information on individual schools is collected through documents and interviews with school principals, teachers, some parents, and pupils/students. Statistics are also obtained from the Central Bureau for Statistics and prevailing studies and evaluations. The program theory is described as follows: “The more the Inspectorate evaluates the compliance of schools with rules and legislation, the more these schools will guarantee a basic level of education, and the more they will live up to these standards, and the more the quality of education is evaluated the more the schools will realize added value to the pupils” (ibid.: 90). Furthermore, the more the Inspectorate carries out compliance studies and school quality evaluation “the more policy makers and politicians will take notice of the Inspectorate's report and will use the results” (ibid.: 91).

These two Dutch evaluation systems both assume that monitoring compliance with rules and regulations and quality controls contributes to improvement of policies, programs, and outcomes. The two program theories conceive of audit and inspection (evaluation) as a rational tool and assume that instrumental use of evaluative information will improve an organization's effectiveness. These theories are also examples of how evaluation is assumed to fulfill an intertwined accountability and improvement function in governance.

Generally, program theory methodology can be applied in at least four ways in the context of governance. The researcher can unpack assumptions about how an evaluation (system) can achieve its intended effects (functions) according to

- democratic governance theories;

- the assumptions of policy makers and evaluation system designers;
- the assumptions of other key actors;
- the researcher's own assumptions.

The first and fourth approaches involve moving beyond the architects' assumptions to consider the functions of evaluation from a governance perspective in which policy makers, system designers, evaluators, and other stakeholders have different roles. In this article two program theories for evaluation in democratic governance, reflecting the accountability and improvement function, are reconstructed according to the first and the fourth approach, matching the state and local or regional model of governance, and the network model of governance respectively.

The unpacked program theories can be helpful in promoting a better understanding of how two key functions of evaluation might work in different models of democratic governance. They can also be used to explore these functions in practice and probe whether the assumptions are valid or not.

<p>1. Initiating an evaluation (system) → raises awareness of current conduct and practice in relation to what is measured/evaluated among accountability holdees → anticipation of continuing evaluation and follow up</p> <p><i>and</i></p> <p>2. Evaluative information/knowledge describes and critically assesses problems, processes, and performance → accountability holdees discuss evaluation findings and current conduct and practice</p> <p><i>and</i></p> <p>3. Administrators and politicians deliberate about policy and program continuation, improvement, or change → politicians decide about program revision, changing action plans if there is a perceived need to do so</p> <p><i>and</i></p> <p>4. Lower-level bureaucrats/ service producers are encouraged/forced to continue or change priorities → they comply with new demands/ implement revised action plans if needed</p> <p><i>and</i></p> <p>5. Follow-up of action taken in response to evaluation between elections → those elected hold administrators/ administration to account and they in turn hold lower levels to account</p> <p><i>and</i></p> <p>6. Citizens use evaluations while voting in election → (re)legitimization of old/ revised or new policies, and the established governance model, or revision of the current governance model</p>
--

Box 1: Assumptions of how evaluation supports democratic governance and key functions in the state and local/regional model of governance

A general program theory for evaluation in democratic governance, reflecting how key functions, mainly the accountability and improvement function, are assumed to work in the state and local/regional model of governance is summarized in Box 1. It illuminates how the intertwined functions of accountability and improvement are assumed to work in a representative democracy governed by national and local governments. It is based on a traditional notion of democratic accountability (Mulgan, 2003) in which citizens and the elected are accountability holders (citizens hold the elected to account, and they in turn hold the administration to account) and governments are accountability holdees (held to account for action taken or not taken). Principal-agent relations follow the delegation of power from citizens via their elected representatives and governments down to street-level bureaucrats and professionals (cf. Behn, 2001; Hanberger, 2009). Basically, this is roughly how actors in representative democracy and the majority of citizens assume evaluation for improvement and accountability should function in democratic governance.

Assumptions 1, 3, and 4 imply that evaluation fulfills an improvement function; assumption 2 a critical appraisal and learning function; and assumptions 5 and 6 an accountability and legitimization function. The evaluation system creates expectations regarding upcoming evaluation; actors should anticipate the next evaluation or follow up, and take action to change the behaviour that the evaluation measures and accounts for. If an evaluation (system) promotes the six “chain-reactions” and functions specified in Box 1 in the real world of governance, it indicates that evaluation serves these models of democratic governance. The program theory also illustrates the interplay of governance and evaluation. The influence of governance on evaluation can become larger if supported by evaluation steering.

Evaluation is not assumed to work in the same way in network governance, mainly because there are many actors in the policy and accountability community who have legitimate roles in governance. Hence, a different program theory must be reconstructed to reflect the ideas of the network model. The assumptions regarding the two functions of evaluation in network governance are set out in Box 2. Assumptions that differ from those in Box 1 are indicated in bold.

As indicated, the assumptions underpinning evaluation in network governance are not quite the same as those in the local and state model of governance because in network governance many actors may act as both accountability holders and holdees. While the program theory for evaluation in the state model is based on a traditional notion of accountability with clear principal-agent relations, assumptions underpinning evaluation in the network model conceive of democratic accountability as conduct that involves the wider accountability environment (Behn, 2001). In network governance evaluation is also assumed to fulfill critical, learning, and legitimizing functions.

If the assumptions of the program theory do not correspond to the real world dynamics, the program theory may need revision. Alternatively, more resources or evaluation steering could facilitate the intended functions.

1. Initiating an evaluation (system) → raises awareness of current conduct and practice in relation to what is measured/evaluated among accountability holdees **and accountability holders** → anticipation of continuing evaluation and follow up
and
2. Evaluative information/knowledge describes and critically assesses problems, processes, and performance → accountability holdees **and accountability holders** discuss evaluation findings and current conduct and practice
and
3. the **accountability environment (administrators, politicians, NGOs, citizens, etc.) deliberate about** policy and program continuation, improvement, or change → **accountability holdees** decide about program revision, changing action plans if there is a perceived need to do so
and
4. Lower-level bureaucrats/ service producers are encouraged/forced to continue or change priorities → **accountability holdees take individual and mutual responsibility to implement revised programs and action plans** if needed
and
5. Follow-up of action taken between elections → **accountability holders (those elected, NGOs, and citizens) hold accountability holdees (national and local governments, administrations, and NGOs/active citizens)** to account.
and
6. Citizens use evaluations **between elections and** while voting in elections → (re)legitimization of old/ revised or new policies and the established **network governance model**, or revision of the governance model

Box 2: Assumptions of how evaluation supports democratic governance and key functions in network governance

Discussion

The framework developed in this article can be used to explore the interplay of governance and evaluation and to unfold the ideal and real functions of evaluation in different models of democratic governance. Three models of governance were used as examples of how governance affects evaluation. There is, however, a need to further explore this interplay in different models of governance (Howlett, 2009; Hansen, 2012; Hertting & Vedung, 2012). The framework presumes that the governance model in which evaluation is embedded influences what kind of evaluation exists in the first place, and that evaluation steering can promote the intended functions of evaluation. Functions of evaluation are explored as effects of evaluative information produced by different evaluations on the one hand, and as evaluation system effects (or constitutive effects) on the other.

The framework can, for example, be used to map out and compare prevailing evaluation (systems) and their functions in the social service or school sector. M&E systems are generally more institutionalized in the governance structure than stand-alone evaluations and could therefore have significant system effects. The ranking of municipalities' elder care, for example, through the Swedish ranking system "Open Comparisons" (SALAR, 2012) is an example of systemic evaluation steering (Hansen, 2008, 2012) and soft governance (Kall, 2010), addressed to many actors in a service field. Research indicates that municipalities tend to prioritize work that is being measured, and some municipalities lower their sights as they consider an average ranking good enough (Lindgren, 2012). It is not known whether "open comparison" or other evaluation systems contribute to improving policy and practice or whether they result in suboptimal public service (see Vakkuri & Meklin, 2006, and Llewellyn & Northcott, 2005, for dysfunctions of evaluation systems). The framework can help to fill the knowledge gap between governance and evaluation in different sectors.

The two program theories for evaluation include five key functions. There is, however, a need to further explore how different functions of evaluation interact in democratic governance. Whether evaluation should play a critical function in democratic governance is not discussed much in the literature. However, a vital democracy needs critical evaluations, and maintaining an established governance structure or model cannot be said to be a fundamental democratic value. Although democratic governance can benefit from critical evaluation, those in power may not finance such evaluations. Hence, there is a need for developing evaluation policies that sustain evaluations which can operate without major restrictions, and there is also a need for independent evaluation agencies to balance the power over evaluation.

How an evaluation (system) describes a policy and policy effects is a political act in a world of symbols, ideologies, and power relations (cf. Furubo & Vestman, 2011; Hanberger, 2003; Stone, 1997; Van Helden et al., 2012; Weiss, 1988, 1999). When an evaluation expert, for example, suggests an evaluation system to match the needs of policy makers, this is a political act that supports a certain model of evaluation and governance. Evaluation as a political practice can also occur when an evaluator or evaluation institution undertakes evaluations for its own purposes (Dahler-Larsen, 2011; Furubo & Vestman, 2011). It is not just the governance model and evaluation steering that may affect evaluation, so may the actors and institutions involved in governance and evaluation.

The program theories for evaluation discussed are ideals. Policy makers and evaluation system designers may have different assumptions, particularly if they seek to use evaluation strategically to implement their own goals. However, few of them will admit to using evaluation to legitimize and substantiate policy (Boswell, 2009), or to using evaluation tactically or symbolically. And you will find no reference to this possibility in official documents. When explored through interviews with policy makers and evaluation system designers, and analysis of official documents, the program theory methodology will primarily

unfold assumptions associated with rationality and statements that are in line with acknowledged norms.

What implications does the framework have for democratic governance, evaluation steering, and evaluation practice? Generally, to maintain democratic governance, evaluations should be designed according to the knowledge and information needs of the existing governance model. To develop democratic governance, evaluation should contribute to the governance discourse by providing narratives on the advantages and disadvantages of existing and new policies, programs, and governance models.

Promoting the accountability and policy improvement function in the state and local-regional model of governance (Box 1) implies designing evaluation to scrutinize formal lines of accountability, focusing on achievement of the state's objectives, and facilitating elite learning. By contrast, strengthening these functions in the network model of governance (Box 2) implies designing evaluation to reflect democratic accountability as it manifests itself in multi-actor policy making, seeking the acceptance of key actors for multiple evaluation criteria that reflect the main stakeholders' knowledge and information needs, and facilitating collective learning.

If evaluation is conceived as embedded in governance, and explored while it interplays with governance, we can arrive at a better understanding of why and how an evaluation is designed the way it is, as well as why a policy or program is implemented and revised in a certain way. In the real world of governance, however, the three models of governance discussed here are mixed, and there are overlapping evaluation systems. Thus there is also a need for more research into how different evaluation systems interplay with governance.

It is hoped that this article will stimulate discussion on the role of evaluation in democratic governance and lead to further exploration of the interplay between governance and evaluation in theory and practice.

References

- Alderson, Charles. & Dianne Wall (1993) 'Does washback exist?' *Applied Linguistics*, 14 (2): 115–129.
- Alkin, Marvin C., Richard Daillak & Peter White, (1979) *Using Evaluations*, Beverly Hills, CA: Sage.
- Behn, Robert (2001) *Rethinking Democratic Accountability*, Washington, D.C., Brookings Inst. Press.
- Boswell, Christina (2009) *The Political Uses of Expert Knowledge: Immigration Policy and Social Research*, Cambridge: Cambridge University Press.
- Bovens, Mark, Paul 't Hart, & Sanneke Kuipers (2006) 'The Politics of Policy Evaluation' in Michael Moran, Martin Rein, & Robert E. Goodin (eds.) *The Oxford Handbook of Public Policy*, Oxford: Oxford University Press.
- Chelmsky, Eleanor. (2006) 'The Purposes of Evaluation in Democratic Society' in: Ian Shaw, Jennifer Greene, & Melvin Mark, (eds.) *The Sage Handbook of Evaluation*, Thousand Oaks, CA: Sage.

- Cousins, J. Bradley & Kenneth A. Leithwood (1986) 'Current empirical research in evaluation utilization', *Review of Educational Research*, 56 (3): 331–364.
- Cousins, J. Bradley, Swee, C. Goh, Clark, Shannon & Lee, Linda E. (2004) 'Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base', *Canadian Journal of Program Evaluation*, 19(2):99–141
- Dahler-Larsen, Peter (2007) 'Constitutive Effects of Performance Indicator Systems' in: Saville Kushner & Nigel Norris (eds.) *Dilemmas of Engagement: Evaluation and the New Public Management*, Amsterdam: Elsevier.
- Dahler-Larsen, Peter (2011) 'Taking One's Own Medicine? The Self-Evaluation of the Danish Evaluation Institute' in: Pearl Eliadis, Jan-Eric Furubo & Steve Jacobs (eds.) *Evaluation: Seeking Truth or Power?* New Brunswick, NJ: Transaction Publishers.
- Furubo, Jan-Eric & Vestman, Ove Karlsson (2011) 'Evaluation: For Public Good or Professional Power' in: Pearl Eliadis, Jan-Eric Furubo & Steve Jacobs (eds.) *Evaluation: Seeking Truth or Power?* New Brunswick, NJ: Transaction Publishers.
- Hanberger, Anders (2001) 'Policy and program evaluation, civil society, and democracy', *American Journal of Evaluation*, 22 (2): 211–228.
- Hanberger, Anders (2003) 'Den dolda utvärderingspolitiken' [Evaluation's hidden politics], *Studies in Educational Policy and Educational Philosophy*, E-tidsskrift 2003:1.
- Hanberger, Anders (2006) 'Evaluation of and for Democracy', *Evaluation*, 12 (1):17–37.
- Hanberger, Anders (2009) 'Democratic accountability in decentralized governance', *Scandinavian Political Studies*, 32 (1): 1–22.
- Hanberger, Anders (2011) 'The real functions of evaluation and response systems,' *Evaluation* 17(4): 327–349.
- Hansen, Hanne Foss (2008) 'Systemisk evalueringsstyring: Potentiale og udfordringer', *Politik*, 11 (1): 6–17.
- Hansen, Hanne Foss (2012) 'Systemic evaluation governance: New logics in the development of organisational fields', *Scandinavian Journal of Public Administration*, 16(3): 48 .
- Hansen, Hanne Foss & Rieper, Olaf (2009) 'The evidence movement: The development and consequences of methodologies in review practice', *Evaluation*, 15 (2): 141–163.
- Hansson, Finn (2006) 'Organizational use of evaluations: Governance and control in research evaluation', *Evaluation*, 12 (2): 159–178.
- Hertting, Nils & Evert Vedung (2012) 'Purposes and criteria in network governance evaluation: How far does standard evaluation vocabulary takes us?' *Evaluation*, 18(1): 27-46.
- House, Ernest R. & Kenneth R. Howe (1999) *Values in Evaluation and Social Research*, Thousand Oaks, CA: Sage.

- Howlett, Michael (2009) 'Governance modes, policy regimes and operational plans: A multi-level nested model of policy instrument choice and policy design', *Policy Sciences*, 42: 73–89.
- Kall, Wendy Maycraft (2010) *The Governance Gap: Central-Local Steering and Mental Health Reform in Britain and Sweden*, Uppsala: Uppsala University, Acta Universitatis Upsaliensis.
- Keane, John (2008) Monitory Democracy. Paper prepared for the ESRC Seminar Series, 'Emergent Publics', The Open University, Milton Keynes, 13–14 March 2008.
- Klijn, Erik-Hans (2008) 'Governance and governance networks in Europe: An assessment of ten years of research on the theme', *Public Management Review*, 10 (4): 505–525.
- Leeuw, Frans (2003) 'Reconstructing program theories: Methods available and problems to be solved', *American Journal of Evaluation* 24 (1): 5–20.
- Leeuw, Frans (2006) 'Managing Evaluations in the Netherlands and Types of Knowledge' in Ray C. Rist & Nicoletta Stame (eds.) *From Studies to Streams: Managing Evaluative Systems*, New Brunswick, NJ: Transaction Publishers.
- Leeuw, Frans & Furubo, Jan-Eric (2008) 'Evaluation systems: What are they and why study them?' *Evaluation*, 14 (2):157–169.
- Lidström, Anders (2004) *Multi-level Governance: The case of Umeå*. Umeå: Umeå University, Department of Political Science.
- Lindensjö, Bo & Lundgren Ulf P. (2000) *Utbildningsreformer och politisk styrning*, Stockholm: HLS förlag.
- Lindgren, Lena (2006) *Utvärderingsmonstret – Kvalitets- och resultatmätning i den offentliga sektorn*, Lund, Studentlitteratur.
- Lindgren, Lena (2011) *Hur används brukarundersökningar? En översikt av forskning och perspektiv på kunskapsöversikter*, Göteborg: FoU Väst rapport 3:2011.
- Lindgren, Lena (i samarbete med Osvaldo Salas och Maria Ottosson) (2012) *Öppna jämförelser – ett styrmedel i tiden eller "hur kunde det bli så här?"* Göteborg: FoUiVäst, Rapport 2:2012.
- Lindvall, Johannes & Bo Rothstein (2006). 'Sweden: The fall of the strong state', *Scandinavian Political Studies* 29(1):47–63.
- Llewellyn, Sue & Deryl Northcott (2005) 'The average hospital', *Accounting, Organizations and Society*, 30 (6):555-83
- Lundgren, Ulf P (2006) 'Political governing and curriculum change – from active to reactive curriculum reforms: The need for a reorientation of curriculum theory', *Studies in Educational Policy and Educational Philosophy*, E-tidskrift 2006:1.
- Liverani, Andrea & Hans Lundgren (2007) 'Evaluation systems in development aid agencies: An analysis of DAC peer reviews 1996–2004', *Evaluation*, 13 (2): 241–256.

- MacDonald, Barry (1976) 'Evaluation and the Control of Education,' in David Archer Tawney, (1976) *Curriculum Evaluation Today: Trends and Implications, Schools Council Research Studies*. London: Macmillan.
- MacDonald, Barry (1978) 'Democracy and Evaluation,' *Public address at the University of Alberta Faculty of Education*, 17 October 1978.
- MacTaggart, Robin (1991) 'When democratic evaluation doesn't seem democratic', *Evaluation Practice*, 12 (1): 9–21.
- Mark, Melvin M. & Gary T. Henry (2004) 'The mechanisms and outcomes of evaluation influence', *Evaluation*, 10 (1): 35–57.
- Montin, Stig (2002) *Moderna kommuner*, Malmö: Liber.
- Mulgan, Richard (2003) *Holding Power to Account: Accountability in Modern Democracies*, Houndsmill: Palgrave/ MacMillan.
- Määttä, Mirja & Kati Rantala (2007) 'The Evaluator as a Critical Interpreter Comparing Evaluations of Multi-Actor Drug Prevention Policy', *Evaluation*, 13(4): 457 – 476
- Ozga, Jenny (2009) 'Governing education through data in England: from regulation to self-evaluation', *Journal of Education Policy*, 24 (2):149–162.
- Patton, M Michael Quinn (1997) *Utilization-focused Evaluation. The New Century Text*, (3rd edition.), Thousand Oaks, CA: Sage.
- Perrin, Burt (1998) 'Effective use and misuse of performance measurement', *American Journal of Evaluation*, 19 (3): 367–379.
- Plottu, Béatrice & Plottu, Eric (2009) 'Approaches to participation in evaluation: Some conditions for implementation', *Evaluation*, 15 (3): 343–359.
- Pollitt, Christopher (1995) 'Justification by works or by faith? Evaluating the new public management', *Evaluation*, 1: 133–154.
- Pollitt, Christopher (2006) 'Performance information for democracy: The missing link?' *Evaluation*, 12 (1): 33–55.
- Power, Michael (1997) *The Audit Society: Rituals of Verification*, Oxford: Oxford University Press.
- Power, Michael (2003) 'Evaluating the audit explosion', *Law & Policy*, 25 (1): 185–202.
- Power, Michael (2007) *Organized Uncertainty: Designing a World of Risk Management*, Oxford: Oxford University Press.
- Rist, Ray C. (2006) 'The 'E' in monitoring and evaluation: Using evaluative knowledge to support a result-based management system' in Ray C. Rist & Nicoletta Stame (eds.) *From Studies to Streams: Managing Evaluative Systems*, New Brunswick, NJ: Transaction Publishers.
- Rist Ray C. & Nicoletta Stame (2006) (eds.) *From Studies to Streams: Managing Evaluative Systems*, New Brunswick, NJ: Transaction Publishers.
- Rhodes, Rod A. W. (1997) *Understanding Governance: Policy networks, Governance, Reflexivity and Accountability*. Buckingham: Open University Press.
- Ryan, Katherine (2004) 'Serving Public Interests in Educational Accountability: Alternative

- Approaches to Evaluation', *American Journal of Evaluation* 25(4): 443–60.
- SALAR (2012) Swedish Association of Local Authorities and Regions http://www.skil.se/vi_arbetar_med/oppnajokforelser (downloaded 2012-08-31).
- Schaumburg-Müller, Henrik (2005) 'Use of aid evaluation from an organizational perspective', *Evaluation*, 11 (2): 207–222.
- Schneider, Anne L. & Helen Ingram (1997) *Policy Design for Democracy*, Lawrence: University of Kansas Press.
- Schwandt, Thomas A. (2002) *Evaluation Practice Reconsidered*, New York: Peter Lang.
- Shulha, Lyn & Bradley J. Cousins (1997) 'Evaluation use: Theory, research and practice since 1986', *Evaluation Practice*, 18: 195–208.
- Segerholm, Christina (2009). "'We are doing well on QAE". The case of Sweden', *Journal of Education Policy*, 24 (2): 195–209.
- Stame, Nicoletta (2006) 'Introduction: Streams of Evaluative Knowledge' in Ray C. Rist & Nicoletta Stame (eds.) *From Studies to Streams: Managing Evaluative Systems*, New Brunswick, NJ: Transaction Publishers.
- Stone, Deborah (1997) *Policy Paradox: The Art of Political Decision Making*, New York: Basic Books.
- Strathern, Marilyn (2000) 'The tyranny of transparency', *British Educational Research Journal*, 26 (3): 309–321.
- Vakkuri, Jarmo & Pentti Meklin (2006) 'Ambiguity in performance measurement: a theoretical approach to organisational uses of performance measurements', *Financial Accountability & Management*, 22(3): 235-250.
- Valovirta, Ville (2002) 'Evaluation utilization as argumentation', *Evaluation*, 8 (1):60–80.
- Van Helden, Jan, Åge Johnsen & Jarmo Vakkuri (2012) "The life-cycle approach to performance management: Implications for public management and evaluation", *Evaluation*, 18(2): 159–175.
- Vedung, Evert (1997) *Public Policy and Program Evaluation*, New Brunswick, NJ: Transaction Publishers.
- Vedung, Evert (2010) 'Four waves of evaluation diffusion', *Evaluation*, 16 (3): 263–277.
- Weber, Edward P. (1999) 'The question of accountability in historical perspective: From Jackson to contemporary grassroots ecosystem management', *Administration & Society*, 31: 451–494.
- Weiss, Carol H. (1977) *Using Research in Public Policy Making*, Lexington, MA: Lexington/Heath.
- Weiss, Carol H. (1979) 'The many meanings of research utilization', *Public Administration Review*, 39: 426–431.
- Weiss, Carol H. (1988) 'Evaluation for decisions: Is anybody there? Does anybody care?' *Evaluation Practice*, 9 (1): 5–19.
- Weiss, Carol H. (1999) 'The Interface between evaluation and public policy', *Evaluation*, 5 (4): 468–486.

Notes

¹ See Bovens, 't Hart & Kuipers (2006) for another approach to illuminating the roles and functions of policy evaluation in the broader politics of public policy making.

² Evaluation is used as a generic concept referring to different kinds of evaluation. The context will tell if the term refers to stand-alone evaluations or M&E-*systems* or both. For the purpose of variation and to remind the reader of the generic use of evaluation I sometimes write M&E-system and evaluation (system).

³ Evaluation steering is recognized as an important instrument in educational governance (Lindensjö & Lundgren 2000; Lundgren, 2006). Christina Segerholm (2009) refers to “governing from behind” when performance measurements of schools and pupils are used for governing, and Jenny Ozga (2009) talks about governing education by data in a consumer-adjusted steering of schools.

⁴ For further discussion of democratic accountability in relation to the three models of decentralized governance, see Hanberger, 2009.

⁵ In Sweden municipal auditors are politicians without seats in the municipal assembly. They in turn commission experts to undertake audits including performance audits.

⁶ The policy and program improvement function is also referred to as the policy improvement function or simply the improvement function.