Evaluation Utilization as Argumentation
Experiences from a Metaevaluation Study on Agency-Level Evaluations in Finland

Author
Ville Valovirta
Department of Political Science
University of Helsinki
P.O.Box 54 (Unioninkatu 37)
00014 Helsingin yliopisto, Finland
tel: +358 9 191 8901; fax: +358 9 191 8832
e-mail: ville.valovirta@helsinki.fi

Abstract
In this paper, an argumentative view on evaluation utilization is being presented. Rather than regarding evaluation information as undisputable knowledge it is viewed as collection of arguments, which can be debated, accepted and disputed. Evaluations consist of three kinds of statements: claims of fact, summative value judgements and formative action recommendations. These claims become objects of individual interpretation, collective argumentation and decision-making in interactional contexts. The argumentative nature of evaluation utilization is illustrated with examples taken from a meta-evaluation study conducted at University of Helsinki. Evaluations of 12 governmental agencies representing different policy sectors and different kinds of evaluation designs were selected for an analysis. The results of the study indicate that in most cases the evaluations have generated concrete actions in agencies under focus and/or within the supervising departments. The explanatory factors for different degrees and types of utilization and subsequently for different roles of evaluations are located in the overall context of the agencies under focus. Two contextual dimensions seem to have had profound effects on the argumentation concerning evaluation results: the degree of pressure for change and the relation between conflict and consensus.
Introduction

Evaluation is an analytical venture based on language. Evaluation process is about questions being asked and answers being given; it is about communicating the results as text and speech. Moreover, the use of the results takes place within an interactional context of discussion, dialogue and negotiations. Contemporary literature on policy-making and administration has paid increasing attention to the argumentative nature of public policy (Majone 1989, Dryzek 1993, Fischer & Forrester 1993). The argumentative nature of evaluation efforts has also been acknowledged by some scholars (e.g. van der Knaap 1995).

However, there have not been any exhaustive attempts to look at how evaluations are actually being used as part of argumentative practices. What there has been are numerous endeavours to find and classify different types of evaluation utilization. This approach does not pay attention enough to the actual argumentative processes in which the evaluations are being discussed and utilised.

In this paper, I will first present a framework for studying argumentation concerning evaluation results. Then I will illustrate the framework with empirical material collected in a meta-evaluation study in Finland 2000. In the study, 12 evaluation studies of governmental agencies representing different policy sectors and different kinds of evaluation designs were selected for analysis. The evaluation processes were examined from the very first initiative to the utilization of their results.

An argumentative view on evaluation

Evaluation is knowledge. When we speak about knowledge, we usually assume, at least tacitly, that knowledge can be right or wrong, true or untrue. But if we look more carefully at what is the essential core of evaluations we will soon discover that it is usually not the factual claims that make an evaluation valuable; it is judgements, assessments, evaluations, in other words giving value for actions and their outcomes; it is also about recommending what actions should be taken in order to improve public actions. The judgements and recommendations are usually the most significant elements of evaluations. And judgements and recommendations cannot be true, but only more or less convincing for particular audiences.

Following Rieke and Sillars (1984) there are basically three kinds of claims presented in evaluations:

1. **Claims of fact.** These claims include descriptions and analytical sights answering e.g. the questions “what has happened?” and “what are the results?”

2. **Claims of evaluation.** These claims are evaluative judgements answering e.g. the questions “does the program work?” and “what are the problems?”

3. **Claims of action.** These claims are formative recommendations answering e.g. the question “what should we do next?”
The reason that I am speaking about claims and statements in spite of facts and evaluations is a pragmatic one. In order to examine how the actual sentences presented in evaluation reports can be used in argumentation we need not take a stance whether the evaluation reports contain truths about the public actions under scrutiny or not. The assumption made here is that in any case they are treated as arguments in practical policy-making. Whether they contain a truth or not, they are anyway considered another contestable arguments in a policy process. Therefore this approach is necessarily also a relativistic one.

Strictly speaking, argumentation is the reasoning that binds two or several claims together in which some claims function as premises and others as conclusions. However, in practical terms, the premises can often be hidden, so that separate statements can be regarded as arguments.

Evaluation is essentially argumentation: making the connection between data and information with the judgements and action proposals. The conclusions might be better warranted than intuitions and individual experiences, yet they can never become undisputable knowledge. Evaluation results can always be debated and disputed. House (1980, 73) puts it strikingly stating, that “evaluation persuades rather than convinces, argues rather than demonstrates, is credible rather than certain, is variably accepted rather than compelling.”

**The argumentative nature of evaluation utilization**

How are evaluation results used in actual policy-making and organisational contexts? I will present a framework for understanding the argumentative nature of evaluation results; to a certain extent, it will even apply to all information processing in interactional settings. The framework is presented in Figure 1.
On individual level the evaluation results are being interpreted and re-evaluated. Actors make interpretations by comparing new information with own experiences, prior knowledge and information needs. According to Weiss and Bucuvalas (1981), people judge the information by two frames of reference: firstly applying a so called “truth test” on which the actor assesses the quality of research and its conformity with prior knowledge and expectations; secondly, applying a “utility test”, in which the feasibility for action and the challenge it poses to current policy are being assessed. The results of this interpretation process are cognitive: they can either confirm the existing comprehensions, generate new or transformed comprehensions, or become refuted.

Policy is not made by isolated individuals, but within organisations and decision-making bodies; policy also evolves in interpersonal and inter-organisational policy networks and advocacy coalitions (Sabatier 1988). The individual interpretations are soon taken under collective deliberation, discussion, and negotiation. The discussions can take place in numerous instances, e.g. executive boards, task forces, cabinet meetings, parliamentary discussions, as well as informal face-to-face dialogues and telephone conversations. The individual interpretations along with acceptances and refutations come under further interpretation, now on a collective level. The practical meaning of evaluation conclusions in a particular context is being generated as a result of interactive processes of arguing, presenting opinions, supporting some viewpoints and opposing others. The basic modes of argumentation (presented in Figure 1) are persuasion, legitimisation, critique and defending. The meaning of
evaluation is being negotiated (Guba & Lincoln 1989) within these argumentational practices.

These interactional processes of collective deliberation and argumentation have resulting effects in three broad categories (following Chelimsky 1997): 1) development, 2) accountability, and 3) knowledge creation. In practical terms, development means decisions and actions; accountability is linked to the legitimacy of actors and organisations, which can become stronger or weaker; knowledge creation refers to changes in thinking, especially as establishment of new shared conceptions and increased awareness of problems and issues.

Since evaluations consist of claims, statements and arguments, it is worth examining how these can be used for different purposes in argumentation. Evaluation statements can become accepted or disputed, criticised or defended (Figure 2).

![Figure 2. The basic argumentative modes](image-url)

Within each of the four sections presented in figure 2, diverse arguments can be presented. It is not only that actors try to convince others that a particular evaluation conclusion is acceptable or not. They also try to give appraisal or criticize, thus increasing or diminishing someone’s legitimacy; and they try to incite or inhibit action.

(A) A positive stance towards a summative judgement can mean:

1. Post hoc legitimising actions already taken using evaluation conclusions (*justificatory argumentation*).
2. Inciting action (other than evaluation recommendations) (*warranting argumentation*).
3. Criticizing something or someone with a negative judgement (*offensive argumentation*).
4. Defending oneself or one’s previous actions using a positive judgement against actual or potential critique (*defensive argumentation*).
(B) A negative stance towards a summative judgement means disputing the judgement on some grounds:
   1. Methodological critique.
   2. Denying or belittling the existence of a problem.
   3. Trying to bypass the judgement and letting it fall into oblivion.

(C) A positive stance towards a formative recommendation means:
   1. Supporting a recommendation.
   2. Trying to persuade others in favour of a recommendation.

(D) A negative stance towards a formative recommendation:
   1. Opposing a recommendation.
   2. Opposing a recommendation and bringing up new alternative courses of action.

I will now illustrate these argumentative possibilities with empirical examples\(^1\) taken from interviews and formal documents collected in a metaevaluation study on agency-level evaluations conducted in Finland.

**Approving a judgment (A)**

There were many examples of post hoc legitimising actions (A1) in the metaevaluation cases.

   “There was some rough twisting about whether the project-based organisation is good or not. And the evaluation group and the subsequent task force regarded the projects to be a natural way to organise research. It was only here finally blessed to be ok. We have been able to refer to this [evaluation].” (Director of Agency A)

This is a major way of showing accountability both to superiors (in this case the ministry) and the public at large. All people and organisations need legitimacy. They need it in order to retain respect and self-esteem and to defend themselves against criticism. They need it also because their superiors make them accountable for their actions. Evaluation can be a very powerful tool for showing good performance for accountability purposes.

Using evaluation judgements one can also speak for other actions than what the evaluators have been recommending (A2). In one case, the agency had been competing for shares of managing an international exchange program. The director needed evidence to show that his agency was capable of taking a bigger share of the program administration. The evaluation – as an independent assessment – served as a powerful instrument for showing the good performance, even though the evaluators had not been discussing the program at all in their report.

   “The ministry decided to increase our share [in the program]. As writing these memoranda I referred often to this [evaluation], that the agency’s customer service has been tested and evaluated and proved competent. There was some market value for it as I anticipated.” (Director of Agency B)

\(^1\) All the interview excerpts are translated from Finnish language by the author.
Evaluation judgements are often critical, putting into question the competence of an organisation or a certain part of it. These critical judgements can be used against organisations or certain parts of them (A3). This will often result in weakened legitimacy of an organisation or its management. In the following example, there had been internal disputes inside the organisation that a reform of the organisation structure had triggered. The evaluation became an instrument in the struggles, which escalated far outside the agency’s own walls. The opposition in house ‘leaked’ the evaluation draft to media in purpose of undermining the trustworthiness of the management. The leading Finnish newspaper headlined:

“International experts conclude: the agency is a haven of bureaucracy and selfishness.”

The result was common turbulence around the institution, where the management had major difficulties in trying to stabilise the working environment. However, this led to a fortified implementation of evaluation results, where the general motivation was to show that rational actions were taken and criticism was being considered thoroughly.

As well as legitimising previous actions one can also defend oneself with a positive judgement against actual or potential critique (A4).

“Our Executive Board was nagging, that our publications are too expensive. The publications were appraised excellent [in the evaluation report] and when the Board started again, I just referred to this. Afterwards the issue has not been discussed any more.” (Director of Agency B)

Refuting a judgement (B)

Not approving the judgements made in the evaluation is of course possible. But in order to argue for the falseness of the evaluation judgements one needs good arguments. One of the most obvious ways is to dispute it with methodological critique (B1). There was one clear example of refuting the evaluation results in the cases examined. The effectiveness analysis of an agency was considered biased by the agency itself. The analysis was therefore left out from the final evaluation report. The agency claimed it biased towards customers with negative experiences of the agency. The ministry, who ordered the evaluation, had formulated the list of possible interviewees. The evaluator had not checked the lists representativeness which caused fierce opposition from the agency’s side.

In case the judgement concerns formulation of a problem, its existence or relevance can be denied or belittled (B2). In one case, after the evaluation report was published the ministry appointed a task force to deliberate on which recommendations were to be implemented. In the memorandum, the task force made a statement for each recommendation considering whether it should be implemented or not. The prevalent mode in the statements was the denying the existence of the problems:

“The implementation of the evaluation recommendation would require changing the power structure between the ministries. There has not been any such problems concerning the management of the [agency] that this kind of action should be taken.”
Another possibility is to try to bypass the judgement and let it fall into oblivion (B3). This is always a viable alternative. For apparent reasons, however, evidence of it is more difficult to collect.

**Supporting a recommendation (C)**

Concerning recommendations, i.e. action claims, the argumentation is somewhat different from arguing about the judgements. Approving a recommendation (a claim of action) can bring forth both passive support for recommendation (C1) or active and forward pushing argumentation (C2). There are several reasons for an actor to approve an action statement. The first one is that it confirms existing comprehensions:

> “The overwhelming majority of these recommendations were issues that had been discussed during last five years; that something should be done. But they were always presented along with some other alternatives. Now that there were clear recommendations by this kind of [evaluation] group, it was much easier from the management’s point of view. There was no need for philosophising whether this or that should be done.” (Director of Agency F)

In light of the metaevaluation cases, this seems to be the general manner of approving the recommendations. Even though it seems to be rare, there were also clear examples of new and fresh insights having made in evaluations:

> “We had strived to reform our organisation piece by piece. But here [in the evaluation report] they recommended a once-and-for-all reform. And that was something we had not hit upon, it was a clear insight.” (Administrative director of Agency G)

Quite often people seem to expect evaluations to offer a lot of ready-made solutions to problems. However, these cases confirm the prevalent conception in evaluation community: evaluations do only seldom offer new, fresh and fruitful insights.

The third possibility is that in spite of the actual content of a recommendation, it is instrumental in showing competent management:

> “Of course we tried to utilise this [evaluation] as much as possible. And act like we’re taking this very seriously. And we did, it did indeed cost something to accomplish this. And above all, to find support for things we were already doing, and find support for our reforms.” (Director of Agency B)

This is apparently overlapping with the second argumentative option, but highlights a clear symbolic value of putting the proposals into action. As Feldman and March have forcefully noted, “information is not simply a basis for action. It is a representation of competence and a real affirmation of social virtue” (1981, 177). Especially in conflict-laden contexts it can become vital to be able to show having implemented all the evaluation recommendations. It can be an effective way to defend oneself against actual or potential critiques.
Opposing a recommendation (D)

In every evaluation context, there are an infinite number of possible reasons for opposing an evaluation recommendation. In the metaevaluation cases there were three major categories discernible for arguing against approving a recommendation (D1). The first one concerns constraints, especially resource constraints:

“One has to be daring and honest and say straight that this just can not be implemented and to present arguments for it. […] A part of these recommendations just couldn’t be carried out in the post-depression circumstances under which Finland was at that time. The resources were very scarce here. Now it would be different.”

The second group of reasons for not taking an evaluation recommendation into account was that it was considered incoherent with other recommendations:

“It was a bit contradictory that they urged to prioritise, that is to discard certain functions in order to strengthen others. At the same time there were proposals of what new functions would have to be created.”

The third group of reasons given was the apparent clashing of proposal with other goals of higher rank. These included regional policy aspects and administrative development ideologies.

“Theyir proposal was against the development line adopted in Finland. At least the Ministry didn’t have the guts to make the Executive Board a purely advisory. The dominant thinking nowadays is that the Executive Board decides certain issues, including budget proposals and annual action plans. We couldn’t make opposing decisions.”

Opposing a recommendation does not necessarily imply inactivity. Bringing up new alternative courses of action and alleging them to better solve the problem (D2), the actor can take an active role in managing the issue.

Arguments against formative judgements (D) give us an outlook on the reasons why evaluation recommendations have not been implemented. The reasons the decision-makers present for opposing evaluation recommendations can reflect real considerations about factors that constrain their implementation. However, a distinction by Majone is highly useful here. He terms “those considerations on which a person acts motives and those which may be used in interpersonal communication reasons” (1989, 29). The important difference is that motives need not always be communicated as reasons and all reasons need not function as the real motives. This makes it reasonable to speak about hidden motives behind presented reasons.

Quite often a conceptual distinction is made between persuasive and instrumental utilization (Leviton & Hughes 1981), or between symbolic and instrumental utilization (Johnson 1998). The difference is said to lie in that the previous “involves interpersonal influence” (Leviton & Hughes 1981, 529) or “occurs when individuals use evaluation information for political self-interest” (Johnson 1998, 94). However, according to the argumentative view on policy-making all significant decision-making involves interpersonal influence. Decisions are not made in a vacuum of individual
philosopher kings but in a cross swell of diverging opinions and interests. Furthermore, distinguishing actions taken on self-interest from actions taken for common good is not an easy task. Discovering the real motives behind actions is usually not possible, and certainly not one that a student of evaluation utilization need to take.

The examples presented above give us a view on what argumentative options there are within diverse stances taken towards evaluation results. Many others would easily be discovered. Yet these examples – so I believe – provide us with a repertoire of the most elementary argumentative approaches available for discussions, negotiations and decision-making.

Utilization of agency-level evaluations in Finland

There has been an outburst of evaluation activity in Finland during the 1990’s. Evaluations have become established as a tool in many government departments and organisations. One major form of evaluation has been agency evaluation in which government agencies have been evaluated in their entirety. Some 30 Finnish central government agencies have now been reviewed after 1990, the total number of agencies amounting to 80.

In a study accomplished during 2000 at University of Helsinki, 12 agency evaluation cases representing different policy sectors and different kinds of evaluation designs were selected for an analysis. The aim of the study was to get a view on how the evaluations had been conducted, what objectives had been set for them and what role they had played in managing and developing the agencies. An important part was to focus on the utilization of evaluations and their learning effects.

The agencies evaluated have mostly been governmental research & development institutes. Some agencies running administrative or service functions have also been evaluated. The governance of the state agencies in Finland is arranged by a “Management by Results” procedure. The ministry gives agencies budget appropriations and defines the tasks they are to accomplish with these resources. The agency is given relatively free hands at using resources in order to fulfil the functions. This can be considered a clear accountability structure (Gray 1997).

Both as a part of structural reforming of the research & development administration and as further impetus given for results-oriented managing of the agencies, institutional evaluations were introduced as a new instrument in the beginning of 1990’s. The evaluation design was mainly adopted from research evaluation field with peer reviews and basic impact analysis of research as the general evaluation design (Ormala 1994). In the majority of the cases, the supervising ministry has invited international experts to evaluate the agency as a peer group. The size of the groups has varied between 2 and 8 members.

Some patterns of agencies’ reactions towards evaluation results have been described above as debating the argumentative modes. The settings on which the discussions have taken place have been numerous: evaluation publication seminars, Executive Boards and Executive Groups, Ministries’ Executive Groups, Results Agreement
negotiations, special Task Forces set up for implementing the evaluations, Parliament, media, as well as all the informal conversations.

The decisions have mostly been made in the following forums:

1) Executive Boards, Executive Groups → agency decisions
2) Results Agreement negotiations → results agreement
3) Ministry’s Executive Group / Secretary General → ministry decisions
4) Government → statutes, legislative proposals
5) Parliament → legislation

The results of the metaevaluation study indicate the following:

- In two cases the evaluation had triggered major organisational reforms.
- In two cases the evaluation had incited moderate organisational reforms.
- In several cases there were changes made in the laws and statutes concerning the tasks and management of the agency.
- In most of the cases there were incremental adjustments made according to the evaluation results.
- In several cases evaluations had incited paying increasing attention to the formulation of a strategy and defining clear priority areas.
- There was only one case out of 12 where the utilisation was clearly neglected.

Even though there were not any reference studies available for making systematic comparisons, it seems that the evaluations have had relatively wide effects on the agencies and their working premises. They have incited concrete decisions and actions, they have been used as tools for showing accountability and they have shaped thinking within the organizations and among their stakeholders. The persons interviewed assessed that the evaluation processes have had important learning effects on the level of agencies.

Since the relevance of the evaluations does not seem to be superior compared to evaluations in average, the explanation for conscientious execution of recommendations has to be sought elsewhere. It seems that the clear accountability relationship with the supervising ministry favours evaluation utilization. Firstly, it is often the ministry who orders the evaluations; secondly, the implementation and reporting of recommendations has been required in results negotiations. The accountability value of utilization is high. Moreover, the general symbolic value of utilization in terms of showing rationality and good management practices is also high.

In explaining the differing outcomes of evaluation utilization, it appeared necessary to take the preliminary context into account. In certain cases the context was very conflict-laden, whereas in others it was mainly consensual. On the other hand, there were evaluations with high expectations of channelling the existing pressures for change, whereas in others the organisation was functioning under stable conditions without any major impediments for change. These two dimensions – conflict vs. consensus and pressures for change vs. stable conditions – are contrasted in Figure 3.
The context defines to a significant extent what argumentative mode becomes prevalent (Sillince 1999). In a consensual context the discussion is probably ascertaining and reconciling so that the basic mode of argumentation becomes development-oriented (cells I and II). Under low pressure for change the evaluation serves mostly as *revising* the existing functions or discovering unknown deficiencies, thus *waking* people up; when the pressures for change are high and visible, the evaluation easily triggers the reforms and functions as *a reform agent*.

On the other hand, in a conflict-laden context the argumentation takes offensive and defensive turns where criticizing, defending, and legitimising is prevalent. The basic mode of argumentation is predominantly offensive so that the evaluation easily becomes a part of disputes and power struggles (cells III and IV). Here the successful performing of evaluation is much more difficult than in the consensual context. The division lines between the opposing parties will easily become present in the evaluation process. At its best, an evaluation can function as *a conciliator* in stabile settings. More probably, it will be used *legitimising* actions and power settings without being utilised as a basis for change. Under high pressures for change it may become *a conflict manager* or *a referee*. In this kind of a case, the risk for producing results that will become accepted only by some parties is high.

**Figure 3.** Explaining the role of evaluation by its context
An interesting feature of evaluation utilization became apparent in evaluation cases where the majority of the recommendations were refuted. Even though the interviewees considered the evaluation unsuccessful due to “wrong kind of visions regarding the future of the agency”, the evaluation process was regarded useful. This can be explained by the argumentative nature of utilization: the fact that the actors had to present good arguments for not taking a recommendation into account forces them to shape more forceful reasons for the omission.

Evaluation offers a possibility of reasoned debate, in which actors have to justify their stances forcefully. This also makes the usefulness of an evaluation less depend on the successfulness of it in terms of the actual acceptance of the recommendations. From this point of view, it becomes extremely important to build up forums of reasoned and open argumentation.

Perhaps then, the most important value of evaluation is that it incites argumentation and directs it to a reasoned debate. When taken seriously, evaluations force people to present well-grounded arguments for refuting the evaluation conclusions and recommendations. This opens up a possibility for new understandings to emerge. It also fortifies the consciousness about the elementary conditions for action.

Conclusions

In this paper, I have first tried to show that the argumentative view on evaluation utilization has relevance in describing the actual process of presenting arguments for and against the evaluation conclusions. I have tried to illustrate the argumentative nature of information use in empirical policy setting with empirical examples.

Secondly, I have presented some general conclusions of the metaevaluation study conducted in Finland in 2000 along with tentative explanations for the relatively wide utilization of the results and the different roles that the evaluations have played in various argumentative contexts.
References


