



LUND UNIVERSITY

Not by the Books: Models, impacts and Quality in 89 Evaluations.

Denvall, Verner

2007

[Link to publication](#)

Citation for published version (APA):

Denvall, V. (2007). *Not by the Books: Models, impacts and Quality in 89 Evaluations.*. Paper presented at AEA.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



LUND
UNIVERSITY

Not by the Books: Models, Impacts and Quality in 89 Evaluations

To be presented at the annual AEA-conference Nov 6 – 10, 2007 in
Baltimore, Maryland, U.S.A.

Associate professor Verner Denvall, School of Social Work, Lund University,
Sweden

Verner.Denvall@soch.lu.se

+46 (0)46-222 94 25

There currently exists a limited understanding of the actual relevance of evaluations and about the work of the evaluator. Despite this limitation, evaluation is an area of research with a scientific and disciplinary development dating back a number of years as a practice carried out by researchers with scientific ambitions, consultants working on assignment, and co-workers within organizations who want to develop their work.

In this paper, an investigation of the current theory, method, and function of evaluation is presented.¹ How are evaluations carried out and whose interests are steering the process? This investigation relates to the scientific discussion about the theory and use of evaluations as well as to the occurrence of specific evaluation models and their relevance.

Purpose and research questions

The purpose of this paper is to present empirical evidence about two elements of the function and procedure of the current program evaluation: do evaluations influence decision-making processes and how are evaluations carried out? Two primary research questions concerning an evaluation are addressed, the first being connected to method and theory and the second to consequences and terms of application.

- The practice of evaluation and evaluation theory are divided into a variety of methods and directions. Here a number of classical issues in evaluation theory are found. Which connections to theory appear, which criteria are utilized, which type of method design appears, and how does the evaluator deal with demands on influence and control?
- Evaluations need to be both effective and legitimate. It is a common perception that the scientific quality of evaluations is insufficient and that evaluations are politicized where carried out. It is unclear how the problems of the organization fluctuate in correlation with the need to do an evaluation and how this can bring about an evaluation. Which relations develop between the evaluator – the orderer of the evaluation and the one being evaluated, and how are the needs of the stakeholders dealt with? Are politicians and public officials affected when evaluations are to be used?

The case

The line of questioning used in this investigation required a researchable application area with a great deal of evaluation activity. There were a number of reasons in favor of performing a case study. Case studies are a recommended method in connection to investigating evaluations since they provide the possibility to closely follow a process and then various case studies can create a basis for generalizations (Fishman 1999). Case studies should be theory-oriented and should utilize numerous parallel sources in order to get massive data access of the context surrounding the case (see Yin 1989).

The evaluation efforts that are now taking place around “the socially-vulnerable residential area” have for a number of reasons been deemed appropriate. Despite the considerable production of scientific knowledge, and in Sweden reoccurring investigations and a number of projects attracting much attention, the modern suburb seems to be nearly resistant to the solutions that have been tested. Knowledge about the consequences of test-measures appears to be scarcely utilized locally. The problem at hand consists of inadequate collaboration between stakeholders,

¹ Founded by the Swedish Council for Working Life and Social Research, http://www.fas.forskning.se/default_206.aspx

weakly developed planning methods, rigid authority structures, rapidly varying problem understanding, late-coming efforts, and merged needs (Healey 1997, SOU 1998:25, OECD 1998).

During the 1990s, a large number of programs were introduced in order to solve these problems in Sweden as well as internationally. The intentions of national and local governments and housing companies have been realized during the early 2000s through extensive interventions and social programs in the form of development contracts and projects for breaking destructive segregation patterns, cleaning up neglected public places or facilitating the path for unemployed people back into the working world.

A Swedish metropolitan initiative has spent approximately US \$700 000 000 in the purpose of reducing social, ethnical and discriminatory segregation and increasing sustainable growth in 7 municipalities and 24 city neighborhoods. The cities and the state have shared the costs.² This program has attracted about 90 evaluators from universities and companies. As an outcome of this almost 100 evaluations have been produced between the years 1999 – 2006.

Approximately 4% of the national budget has been spent on evaluation (80 million SEK) and it is unknown how much has been contributed by cities or other public authorities. In the national investigation “Storstad i rörelse”, (SOU 2005:29), translated as “Metropolis in motion”, the quality of evaluations is reviewed. This review was deemed necessary because the task of the evaluation was to compile knowledge about the work of big cities based on the results of the evaluations. A large amount of reports were deemed to be lacking in quality, something which runs the risk of “reinforcing a skeptical attitude between practice and research” (aa:104).

The municipalities have carried out a large number of activities, the exact amount of which is not known. This is not clear due to reasons such as co-financing with other actors, unclear forms of reporting, and different definitions of what an activity is. These activities have been in the form of different projects for dealing with social exclusion and for strengthening the labor market in segregated residential areas. An example of this can be seen in the municipality of Haninge outside of Stockholm, where 10% of the government subsidies have ended up and where more than 100 activities have been carried out. In every municipality a number of contributory stakeholders were present with varying demands and expectations about the activities as well as the evaluations. Internal and external evaluators were appointed by the municipalities at hand, which lead to a large variation in evaluators and evaluation methods. The case at hand is considered to well reflect the current conditions of evaluations that earn a lot of political attention, complex processes, and demands on control and participation from stakeholders (see Guba & Lincoln 1989).

Three sub-studies have been carried out in this investigation; document reviewing, interviews with evaluators, and a municipality-level survey. The *document reviewing* consists of all 89 evaluations which have been produced by evaluators engaged by municipalities and by public authorities from the year 2000 up to and including the spring of 2005. The delimitation has been set to evaluations from the Swedish Integration Board that had a direct task of supporting municipalities and carried out some basic knowledge-compilations, i.e. statistics. Research reports and school papers are not included, as the studies should have the aim of evaluating activities within the program. A template for the review was created and a number of reports have been reviewed by two individuals each. The *interviews with evaluators* were done via telephone on the basis of a fixed questioning guide which was sent out in advance during 2004/2005 and has included each active evaluator, that is 94 people of which 6 evaluators were not able to be reached

² <http://www.storstad.gov.se/>

or declined the offer to participate. The *municipality-level survey* took place during the summer/fall of 2003 at which time 600 surveys were sent via mail to project and activity leaders, centrally concerned public officials and politicians in seven municipalities and 24 city districts. The total number of surveys varied per municipality and was chosen in cooperation with the respective municipality coordinator. The answer frequency was 66%, equal to 398 answers. Falling off amongst politicians occurred more extensively than among public officials.

Evaluation problems related to this investigation

What is considered an evaluation depends on who is speaking – researchers from different disciplines, national authorities, orderers of evaluations, or other stakeholders. In this investigation an evaluation is considered a systematic ex-post evaluation of public programs (see Vedung 1997, Åberg 1997).

Evaluation research is hardly considered a discipline. Since it is carried out in close contact with the surrounding society it has been deemed transdisciplinary (Scriven 1998, 1991). Its production of knowledge occurs in a natural environment, often together with and under strong influence from stakeholders. The results are to be applied and practiced within organizations and political areas. The demands of stakeholders on evaluators and the expectations about what an evaluation should result in vary significantly. This also holds true for the opinion of evaluation research concerning the right of interested parties to be influential. Evaluations are carried out as practical academic research – evaluation research – by externally engaged evaluators, but are also carried out as internal tasks of public officials.

In Sweden, evaluation research is associated primarily with the research traditions of pedagogy (Dahlöf 1989, Franke Wikberg 1997), political science (Vedung 1997, Premfors 1989), and business economics (Rombach & Sahlin-Andersson 1995). A number of significant activities occurred in the beginning of the 1980s (Nilstun & Hermerén 1994, Eliasson & Nygren 1980). More recently, the problem of criteria has been investigated (Karlsson 1995) and Åberg (1997) demonstrates how evaluations must satisfy both rational and legitimacy-strengthening requirements.

There is limited understanding about how the evaluation needs of organizations and political areas are connected to the actual evaluation planning. Concerning certain government sectors the picture is quite clear; however it becomes a bit fuzzy regarding municipal activities. (see Vedung, Furubo & Sandahl 2000). Nowadays researchers are being affiliated with interest organizations or to authorities where they attempt to create an evaluation discourse (Tengvald 1995, Soydan 1998, Karlsson 1999). The last decennium has seen the establishment of an evaluation department within a majority of authorities and departments.

Evidence has now been brought forth as a prominent issue for evaluation and is linked to experimental designs (Socialstyrelsen, National Board of Health and Welfare 2000). A deep rift between advocates for the concept and those with reservations does occur (see Davies et al 2000 for a review of the arguments between champs and critics). Evidence-based practices can be applied within a variety of fields: social work, medicine, education, etc. An evidence-based policy has come to be seen as an urgent concept in the modernization of public organizations (Davies et al 2000). The evidence-based practice is also said to have developed into a world-wide phenomenon without national borders and “What works” is a common catchphrase (Trinder 2000).

In the Swedish context things that are still lacking are knowledge summaries, studies of ongoing evaluations, and assessments of the relevance of different methods and of the function and use of evaluations, i.e. meta-evaluation. One could believe that the need for and availability of evaluations varies from one political arena to another. Within professionalized fields such as health and medical services, evaluation (evidence-based) is a more and more self-evident element. However, it is unclear what impact that demands on performing an evaluation have within the more clearly politicized arenas. Two central issues – problems if you will – that affect evaluations are firstly the method and theory issue – how evaluators work, and secondly the issue of legitimacy, i.e. what is the value of an evaluation for concerned public officials and politicians.

The first problem is linked to the theoretical and methodological position of the evaluators. Evaluation research does not have a homogenous core, but is rather divided when it comes to the connection to theory, choice of criteria, methods and the significance of the relationship to stakeholders (Shadish & Cook & Leviton 1991, Weiss 1998, Fishman 1999, Alkin 2004). Today, evaluations are characterized as being complex and needing to balance a multitude of needs concerning control, methods and theory awareness, and the involvement of stakeholders (Guba & Lincoln 1989). It is unclear which theoretical and methodological directions that the practice of evaluation has entered upon and if this can be derived from definite problems or demands of stakeholders, especially in a Swedish context. The discussion about the usefulness of evaluations has been a thriving topic, especially in the U.S.A. and the need for a common theory about the influence of evaluations within different sectors has been emphasized (Kirkhart 2001). This study links to similar issues which also have arisen in the international evaluation discourse (see Mark 2004).

Problem number two has been investigated within research about the usage of knowledge. Research knowledge and surely even evaluation knowledge was for a long time regarded as useful if it could support administrative and political decision-making, thereby contributing to problem-solving (Nilsson & Sunesson 1988, Weiss 1980, Vedung 1995). At that time this form of knowledge constituted a natural and desirable foundation for decision-making and researchers should be involved in order to connect the goal to the means. Research could even be utilized afterwards in order to gain knowledge of the results and effects of an activity. Through this perspective research enriches the political work as well as that the level of professionalism increases by the addition of expert knowledge.

Since the 1950s, international social science research has within a majority of disciplines condemned this optimistic picture of research and science as engineering (Simon 1945, Lindblom 1959, Weiss & Bucuvalas 1977, Wildavsky 1979). The Swedish researchers Kjell Nilsson and Sune Sunesson were able to demonstrate an instrumental use of research in only 20% of the investigated cases of social activities (1988). Interest has therefore been aimed at discerning if alternative types of usage are available. A large number of models for actual application have thus been suggested; currently “impact” is recommended instead, i.e. to what degree knowledge from evaluations has any sort of influence (Kirkhart 2001, Henry & Mark 2003).

The issue of theory and method

Evaluators need to apply some sort of systematic methodology. In this attempt, some of the evaluation practices lean heavily on scientific research discourse whereas others try to advocate reflection and self-valuation in daily life (Eliasson et al 1990). This is described as an application of positivistic respectively post-modernistic scientific understanding (Fishman 1999).

In one summary, no less than 22 different evaluation models are presented (Stufflebeam 2001). Stufflebeam's conclusions are that none of these can be proclaimed to be the best; rather, all methods have different benefits and drawbacks. However, Stufflebeam argues that only nine of these are adequate to be used as professional tools. Irregardless of ones opinions about his conclusions, the ambition to identify evaluation models is an important step for discerning the problems and possibilities as well as the knowledge-theoretical distinctions that are attached to different concepts.

It is conceivable that evaluators carefully observe some form of practice which is linked to a fixed theoretical and methodological point of departure. Those who open one of the many guidelines that are available about evaluation would easily get the impression that the given advice is taken into consideration in practice (Krogstrup 2003, Fitzpatrick, Sanders & Worthen 2004, Rossi, Lipsky & Freeman 2004). Those 22 models that Stufflebeam (above) has gone through also have a number of fixed practices (methods) in tow. There is no lack of handbooks, but where we have limited knowledge is in how applicable they are in practice. It is conceivable that evaluators instead do not care very much about theoretical and methodological concepts, but rather have a more pragmatic relationship to evaluations in practice.

A series of viewpoints on evaluation exist where different types of evaluations are categorized in the form of models. The term model must be linked to a theoretical starting point; models mirror internal variations and explain them, and can be linked to practical methods and tools. The model is in its foundation an analytical simplification where a number of elements are placed together. The model's capacity is determined by how clear it can explain variations and how these variations fit within the model. Sometimes the term "perspective" is used instead of model. As I understand it, this term is used without much precision and concerns a general attitude with a broader starting point. One element of a model is that it clarifies under which preconditions and with which methods it should be used. Models are often not given since a clear understanding about the origin and criteria for making borders can be missing. Moreover, models are often added to an original version during the course of a number of years; models become a "complex dialectic interplay between text and context" (Krogstrup 2003:55).

Theories constitute a collection of terms and empirically secured agreements about how the world should be described, i.e. the parts of the world which we have knowledge about; or rather, what some advocates want to claim to be the correct way to summarize and understand reality. Theories constitute a type of discursive standpoints and formations that can and should be studied based on the claims they place on how our surroundings should relate to them. In one outline evaluation theories are categorized into three areas (methods, value criteria, and use) (Alkin 2004). This same author has also asked different theoretical advocates to comment on their work and their position in relation to other evaluation theorists. The book is in many ways an impressive piece of work, but it says nothing about how theories are used in practical evaluation work or what value theory has for the evaluator. However, large parts can be incorporated into how evaluation is carried out. Maybe one could speak of the element of theory where a selective process takes place and certain parts are applied while others are left out. The use of theory has been investigated in other studies although hypothetically and not empirically (Christie 2003, Alkin & Christie 2005, Christie 2007).

Operationalizing of evaluation directions

The principles for how evaluations are sorted vary immensely. In Scaninavian evaluation research suggestions have been made lately that sorting should be based on the value-criteria of democracy

(Vedung 2002), critical realism (Morén & Blom (2003) or their policy implications (Krogstrup (2003). There are good reasons to start from two main directions. The first focuses on result oriented and effect oriented directions and is utilized in particular for assessing activities. It has the evaluation form often referred to as summative or assessment and control oriented. A second evaluation form is formative and contributes to improvement and method development. These types are understood to have a potential in development work where they can increase the understanding of often complex realities.

This investigation has taken its starting point in this rough way of sorting; partially wanting to find out what the orderers and users of evaluations in municipalities have wanted the evaluations to focus on and partially investigating via review of texts (evaluation reports) and via questioning evaluators about what direction evaluation work has had. In this case we have distinguished between evaluations that have put emphasis on:

| | |
|---|---|
| A. Control | A broad range of evaluations investigating outcome and empirical support of interventions. Randomized control trials (RCT) have the highest status. |
| B. Process, learning and quality development of projects and activities | These evaluations are often open, process-focused and an interactive form where the criteria originate successively, often descriptive and in cooperation with those affected. |
| C. Users' experiences | Can be evaluations that strive to create dialogue and influence of those affected and that put a high value on aspects of democracy. But also in the form of quality measurements where the user's experience of service should be measured. |
| D. Democracy and power | Evaluations with a stated position for improvement through increased influence and learning for those affected, especially weak groups. Almost a consultant/advocate role for the evaluator where those affected are engaged and define criteria. |
| E. Stakeholders | Evaluations that are open and negotiable with focus on creating understanding for the program's goal and guarantees a high level of usage amongst stakeholders. |

Public officials/politicians and evaluators have answered the same questions. In this summary that is presented below, the answers from public officials and politicians have not been separated. Representatives from the municipality were asked to take a stance to the question: What in your opinion is the most important thing an evaluation can contribute with? And the evaluators: Which direction has evaluation work had? The specific questions were linked to the five models that I presented above. The answer choices "Not important" and "Limited importance" have been combined into "Less important" in Table 1, the same has been done for "More important".

Table 1: Attitudes towards evaluation work of public officials/politicians in seven municipalities as well as their evaluators. Percent.

| Model | Survey question | Pub. off./pol N = 396 | | Evaluators N = 88 | |
|------------|--|-----------------------|----------------|-------------------|----------------|
| | | Less important | More important | Less important | More important |
| A. Control | To value the outcome of the activities | 3.2 | 96.8 | 24.7 | 75.3 |
| B. Process | Develop learning amongst personnel | 9.6 | 90.4 | 21.8 | 78.2 |
| | Contribute to quality | 4.5 | 95.5 | 11.8 | 88.2 |

| | | | | | |
|-----------------|--|------|------|------|------|
| | development | | | | |
| C. User | Investigate users' experiences | 8.6 | 91.4 | 25.2 | 74.8 |
| D. Democracy | Contribute to increased influence amongst neglected groups | 14.7 | 83.3 | 38.3 | 61.7 |
| E. Stakeholders | Give support to collaboration | 17.8 | 82.2 | 31.6 | 68.4 |

Source: Survey to public officials and politicians in seven municipalities and interviews with 88 evaluators

Representatives of the municipalities were keen to see that everything should be brought to light in connection with the evaluations. They often gave answers falling into the category “more important” and suggested that the evaluations contributed to increased knowledge about results as well as support to designing projects. This was in contrast to the evaluators who were forced to prioritize amongst the wishes of those ordering the evaluation when they were carrying out their evaluations. Some interesting patterns can be distinguished.

There is a clear dividing line between the expectations of public officials and politicians regarding that the evaluations are to provide knowledge about effects and outcome and in how the evaluators have answered to these expectations. To value outcomes of activities is seen as the most important aspect amongst the municipality representatives and has been indicated by nearly each and every one of them who was questioned. This is contrary to the evaluators where one in four has judged it as a less important issue to address. The issue about how effects best should be measured is a hot topic in evaluation circles and has led to much debated evidence hierarchies based on assessments of the positive and negative aspects of each method (Qureshi 2004). According to a more orthodox line of thinking, only experiments with randomized control groups can keep extraneous factors under control. None of the 89 examined reports has carried out such a classic effect study with a control group.

Table 2: Degree of evidence in 89 evaluations from Metropolitan Initiative published between 2000 – 2005. Number of evaluations.

| Evaluation form | Number of evaluations |
|-------------------------------|-----------------------|
| Reviews | 0 |
| Randomized controlled studies | 0 |
| Quasi-experimental studies | 1 |
| Pre – post design | 6 |
| Expert report | 82 |

Source: Evidence ladder from the Centre for Evidence-Based Social Services (CEBSS) and 89 evaluation reports.

Based on strict inclusion criteria that are used by for example Campbell Collaboration or the government agency the Swedish Council on Technology Assessment in Health Care (www.sbu.se), none of the published evaluations would have been deemed of high enough quality to be acceptable. It should be stated that there are considerable differences between the direction of SBU and governmental and municipal orderers of evaluations concerning big city politics, however there is a substantial discrepancy between the requirements of the quality of the foundations amongst different orderers and evaluators.³

Table 1 further demonstrates how municipality representatives as well as evaluators put a high importance on process-related elements. The evaluation agreement between the orderer and the evaluator shows that there is a clear desire from the orderer to quickly want to put the evaluation

³ SBU is an agency that compiles knowledge and gives clinical recommendations based on evaluations.

results into improving the projects. In a number of cases the evaluator has worked almost consultatively (see below). It also is seen that issues regarding service users and influence are not seen as especially pressing. This is noteworthy since a prioritized policy has been the bottom-up perspective and to create the possibility for neglected groups to have access to decision-making and planning arenas (Edström & Plisch 2005). It seems as if democracy and empowerment perspectives have yet to get a clear form within Swedish evaluation unlike the international evaluation discourse (Fetterman 2001, House & Howe 1999).

Table 1 was based on the statements of the evaluators about their own work. What are the actual evaluation methods as shown in the reports?

Table 3: Evaluation methods in 89 evaluations from Metropolitan Initiative published between 2000 – 2005. Percent.

| | | |
|-------------------|---------------|------|
| Evaluation method | Process | 19 % |
| | Descriptive | 21 |
| | Goal | 16 |
| | Mixes 2 | 23 |
| | Mixes 3 | 5 |
| | Miscellaneous | 16 |

Source: 89 evaluation reports

The evaluations have been assessed based on the aim that the authors have demonstrated in the text, complemented with assessments based on planning and presented directions. More than one third of the evaluations mix different directions, such as process-related purposes with wanting to evaluate according to fixed goals. In the category “miscellaneous” evaluations with directions such as purely theoretical, quasi-experimental or explanatory are placed. Especially notable is that one fifth of all evaluations have a descriptive, non-valuating purpose. This high proportion of process-related purposes is reflected in the scientific ways of working that have been used.

Table 4: The occurrence of quantitative and qualitative methods in 89 evaluations from Metropolitan Initiative published between 2000 – 2005.

| Quantitative method | Mixed | Qualitative method |
|---------------------|-------|--------------------|
| 7 | 10 | 72 |

Source: 89 evaluation reports

Four out of five evaluators uses only qualitative methods, preferably interviews and document studies. Only one tenth rely completely on quantitative methods. Often times the coverage of methods is very meager; on average 3.5% of the total amount of pages is used for this purpose. This leads to big problems regarding validity and reliability if the reader does not get to know how the interviews have taken place, selection procedures, and based on which procedures the data has been analyzed. Naturally, there are large variations with a number of exceptions where report authors carefully give an account of the different conclusions that can be drawn and which limitations exist. The primary impression is that evaluation authors have not only taken the easy way out when it comes to presenting the method, but have also avoided self-critically discussing the range of empirical evidence and grounds for valuations. This leads us to a final table where a number of aspects central to evaluation are gathered.

Table 5: The occurrence of assessment criteria, critical approach, linkage to evaluation discourse, as well as suggestions for change in 89 evaluations published between 2000 – 2005. Percent.

| | | |
|---------------------|--------------------------|------|
| Assessment criteria | Missing completely | 36 % |
| | To a certain extent | 34 |
| | To a considerable extent | 30 |
| Critical approach | Missing completely | 36 % |

| | | |
|---------------------------------|-------------------------------|------|
| | To a certain extent | 36 |
| | To a considerable extent | 28 |
| Linkage to evaluation discourse | Missing completely | 74 % |
| | Certain, - 2 | 21 |
| | Considerable, 3 - | 6 |
| Suggestions for change | Missing completely | 69 % |
| | To a certain extent, - 5 | 20 |
| | To a considerable extent, 6 - | 11 |

Source: 89 evaluation reports

The issue regarding the occurrence of assessment criteria is central in connection to evaluation. Since making some sort of judgment is an inevitable part of evaluation it is essential that these bases are presented. Evert Vedung (2002) differentiates between substance models and economic models. Substance models are evaluations where assessments in one way or another are focused on the activity in focus but do not take costs into consideration, which occurs in the economic models. An additional analysis level is based on to what extent the models are descriptive and collect assessment criteria internally from the activities or to what degree that they are prescriptive and are assessed based on external theories, goals or ideologies. Approximately one third of all evaluations lack some form of presentation of grounds for assessment, around one third discuss it to some degree and in the remaining third this is accounted for clearly, i.e. in the form of theories to which the studied activities are assessed or as regards to the goals of the activity or the government.

Closely linked to the criteria aspect is whether the evaluator has a critical, valuating approach or mostly a descriptive purpose. Probably the clearest advocates that evaluators should take a position to this are Ernest House & Kenneth Howe (1999). Of the idea that the evaluator can be devoid of value judgments they formulate a problem. The authors suggest that evaluators cannot choose values themselves, rather that this should be done in a process together with those being evaluated. An actual evaluation situation is complex and is dependent on a number of practical considerations. But this also has serious limitations; an evaluator can not take all material into consideration but rather must set boundaries and choose the way to proceed. House and Howe argue that evaluators can very well draw objective conclusions by reporting about their analysis and by following a professional direction. Evaluators should, according to their own approach, relate to theory and democracy. We have already noted that approximately one out of five evaluations has a descriptive purpose, which is reflected in the reports. Approximately one third of the evaluation reports take a clear stance to the activity it has assessed.

An obvious lack is in regards to how the evaluation reports refer to and link to the Swedish and international literature on evaluations. One could expect that this occurs by discussing method issues and that an argumentation for a special direction through referencing how others have done or that results are discussed in light of other evaluation reports. Three of four evaluation reports completely lack such a connection except to other evaluations carried out within the Metropolitan Initiative. This can be connected to the fact that 64 of 88 (73%) evaluators stated that they lack education in evaluation which can be a contributing reason as to why they are not familiar with this discourse.

To conclude, the reports were assessed based on the occurrence of suggestions for improvement and policy implications. Seven of 10 reports completely lack such suggestions.

In summary, a picture appears of an evaluation practice that is primarily qualitative in nature. It is process or descriptively focused and concentrated on contributing to process and improvement support for the projects that are studied. The investigations are more activity-oriented than user-

oriented. The reports are often times poor at providing discussion about method as well as at giving suggestions for improvements, and the starting points and valuations that possible conclusions and assessments are based on are rarely accounted for. Effects are studied primarily with qualitative methods. The purchaser and the orderer have many demands as to what the evaluations should be able to contribute to and are seen as having difficulty in providing clear instructions for orders.

The issues of legitimacy and knowledge

There are planning researchers who are currently of the opinion that *ex ante* inquiries should be replaced by ex post-oriented actor-based and dialogue-based evaluations (Huw & Healey 1991, Healey 1997). To plan is looked upon as being outdated and belonging to olden times – to evaluate seems more modern and where ones efforts should be spent. According to this point of view, evaluations are expected not only to deliver knowledge back to the decision-maker about the results of the activities, but also in some way influence the reality which they assess. Evaluation as a practical activity can be described as an issue of legitimacy where the value of rationalized investigation methods and carefully formulated decision bases are assessed in a political context. The relationship between science and politics is a timeless issue within evaluation (Karlsson Vestman & Connor 2006).

Research about knowledge use has often a questioning attitude toward instrumentalism and rational decision-making. It is not the possible value of the research results per se (for example topicality, news value, scientific support, reporting style) that decides if they are to be used or not. Usage can however have its origin in clashes of interest and changing power relations since investment in knowledge can be a way to support certain interests within an organization or the organization’s relationships to its surroundings (Nilsson 1992). Likewise, research has a conceptualizing function due to its informative effect: one reconsiders and gets new ideas. The relationship between research and the user can be describes as a complicated exchange relationship where advocates for scientific knowledge become subject to being influenced by actors on different levels in different structures with varying interests, something that even leads to controversy (Brante 1984). This occurs not least due to the ambition of professional groups toward increased legitimacy as they want to increase their connection to research and through the alliances that are created between research and user.

Since the 1970s, international discussions about knowledge use have been strongly affected by Carol Weiss’s research. Her description of how knowledge “creeps” (1980:397) up on the user had a huge impact at first. The image of this passive user who slowly reacts to and takes in research has in recent times been complemented with an image of a user who is active and aware. According to this view, knowledge is utilized based on fixed points of interest and value-based standpoints. Scientific knowledge constitutes in this case one of many means for influencing an intended development (Patton 1977).

Let us return now to our case study. Have public officials and politicians read the reports, have the evaluations had an impact, and what knowledge do the evaluators have about the fate of their reports?

Table 6: The amount of public officials and politicians in seven big city municipalities that have read an evaluation regarding their own municipality. Percent.

| | Total |
|--------------|--------|
| No | 24.2 % |
| 1 evaluation | 19.0 |

| | |
|-------------|------|
| 2 – 3 | 25.0 |
| 4 or more | 26.3 |
| Do not know | 5.5 |

Source: Municipal survey 2003, N=384

Approximately one fourth of those questioned have not read any reports and just over one fourth are well informed and have read four or more reports.⁴ How has this happened?

Table 7: In what way public officials and politicians in seven big city municipalities have come across evaluations regarding their own municipality. Percent.

| | |
|------------------------------|-------|
| | Total |
| Verbally by the evaluator? | 20 % |
| Verbally at a staff meeting? | 5 |
| Read the summary | 15 |
| Read the entire report | 21 |
| No answer | 38 |

Source: Municipal survey 2003, N=396

A relatively large percentage of those who were asked have avoided answering the question; the majority of these can presumably not have read any reports. The result should however please the evaluation authors as well as the municipal evaluation orderers as it shows that the reports are read to a relatively high extent.

Now, have the reports contributed to increased knowledge? How are the contacts with the evaluators perceived? The respondents from the seven municipalities as well as the evaluators have been asked about this. The question has been asked as the statements about to what degree one agrees or not. The possible answers “Yes, absolutely” and “Yes, partly” have been combined to “Predominately yes”, and the same has occurred for “Predominately no”. Discrepancies between the possible answers in the respective categories can be attributed to the “Do not know” alternative. Falling off varies between 7-10% per question. In Table 8 below nine questions are presented.

Table 8: Attitudes toward evaluation work amongst municipality representatives from seven municipalities as well as their evaluators. Percent.

| | Municipality representatives N = 396 | | Evaluators N = 88 | |
|--|--------------------------------------|------------------|-------------------|------------------|
| | Predominately yes | Predominately no | Predominately yes | Predominately no |
| The evaluations give public officials new knowledge | 86.1 | 7.9 | 90.7 | 9.3 |
| Public officials have had good support in their work by meeting the evaluators | 43.3 | 37.4 | 84.9 | 11.1 |
| The evaluators are perceived to be too invisible | 44.2 | 27.9 | 29.1 | 65.1 |
| Public officials appreciate | 84.7 | 3.4 | 73.2 | 22.0 |

⁴ As a control question the respondents were asked to name the title of the most recent report they had read.

| | | | | |
|---|------|------|------|------|
| the critical thinking that the evaluations represent | | | | |
| The evaluations are too distanced from the daily realities | 45.5 | 41.7 | 41.9 | 56.9 |
| The evaluations have been made available to the activity at a good time and in a good way | 35.3 | 36.7 | 57.0 | 40.7 |
| The evaluations give public officials good decision support | 58.5 | 21.4 | 63.6 | 25.9 |
| The evaluations have been a useful instrument in the planning of new projects | 38.2 | 36.5 | 47.7 | 29.0 |
| The evaluations have been useful in assessing if projects are to continue | 50.0 | 29.2 | 60.7 | 26.2 |

Source: Municipal survey as well as interviews with evaluators

The tables show that possible apprehension about whether or not the evaluations would contribute with new knowledge is premature. On the contrary, it is somewhat surprising that as many as half of the respondents from the municipalities consider the evaluations to be useful as instruments for assessing projects and that evaluations provide good decision-making support. The question remains, is the knowledge really put to use? This is followed up below. The views of the evaluators and municipality representatives are in agreement with a few exceptions. The evaluators clearly overestimate the importance for public officials in meeting with them; they have instead been perceived as invisible to a rather high extent. The timing of the evaluations in relation to the needs of the activities has not been entirely good. Evaluators feel that they deliver the reports in time while the municipality representatives are not as convinced. Here there is a large number of “Do not know” answers, likely due to that one has not taken part in or is familiar with this relationship.

Municipality representatives as well as evaluators have been able to take a stand regarding what consequences that the evaluations have had for the evaluated activities. They have been asked to relate this to the most recent evaluation they have studied the contents of and have taken part in. Fall out from this type of question is naturally high since many do not have sufficient knowledge in order to express themselves in this area.

Table 9: The occurrence of knowledge about consequences of evaluations for evaluated activities within the Metropolitan Initiative. Percent.

| | Public officials and politicians N = 396 | Evaluators N = 88 |
|--|--|-------------------|
| No significance to speak of | 22,2 % | 15,9 |
| Resulted in that the activity was made permanent | 11,1 | 6,8 |
| Resulted in that the activity continued as usual | 9,3 | 6,8 |
| Resulted in a minor change in the activity | 8,3 | 18,2 |
| Resulted in a major change in the activity | 3,8 | 8,0 |
| Resulted in that the activity was | 2,0 | 2,0 |

| | | |
|----------------|------|------|
| discontinued | | |
| Answer missing | 43,2 | 42,1 |

Source: Municipal survey as well as interviews with evaluators

It can be seen as somewhat surprising in that as many as four of 10 evaluators do not know the outcomes from their evaluations. Two types of explanations to this have been presented. On one hand, some evaluators indicate that it could be too early to comment since they in many cases were in the process of reporting or had recently finished a report. Alternatively, they had finished their task and had not had continued contact with the orderer or that this contact was taken care of by other representatives for the evaluation group.

Around one third (34.5%) of the questioned municipality representatives state that knowledge from evaluations has influenced the evaluated activities. A total of 41.8% of the evaluators have reported the same experience. Only on a small number of occasions have the evaluations been used as a basis for discontinuing an activity. For many of those questioned, it has been difficult to assess if the evaluations are used; there are also large differences in the meaning of the word “use”:

“I don’t know if the evaluation has had any significance but through a number of different reports and articles in various internal periodicals it has contributed to that the activities can continue as part of the normal activities but in a different form.” Public official

“It’s difficult to know what is being asked for here. Is it the evaluation of my own project or other projects? Evaluation of my own project has been very important for me. It has led to control of my own work. However, I haven’t seen that much of other people’s evaluations.” Public official

Many clear attitudes exist stating that the evaluations come at a late stage in the process:

“A number of different evaluations in different phases of the process have occurred, many people and questions but nothing has any particular effect on the projects since the result of the evaluations have come altogether too late or not come at all” Public official

“The evaluations come about too late in order to have any real effect but they can result in that those people working in the project develop their way of working. The evaluations often have broader perspectives than one specific activity.” Public official

That someone would develop their way of working is a type of result, an “impact”. Many who participated in the survey refer especially to the importance that individual evaluations have had on investments in language: *“The evaluation resulted even in that the government has decided to incorporate bilingual teaching in compulsory school grades 7-9 for a four-year trial period for students with a mother tongue other than Swedish.” Public official*

Another aspect that is pointed out in many of the evaluations (Bunar 2004, Jensen 2004) is the importance of municipalities organizing the receiving of the evaluations in an active way. These attitudes are also found among the respondents: *“An organizational development toward a learning organization is needed in order to be able to reap the benefits of the evaluations.”* Respectively: *“We are way too bad at following up evaluations and to use the outcomes from it in the programs.”*

In conclusion, this investigation does not differ very much from others that have investigated knowledge use. It is possible that the results be interpreted that knowledge gained from evaluations is used instrumentally (34.5%) to a larger degree than in the investigation by Nilsson & Sunesson from 1988 that reported 20% use in research. Since this is dealing with evaluations that are possible to apply directly to activities, it is not particularly surprising that the amount of

applicable knowledge is somewhat higher. It does not seem as though this has a connection to the scientific quality of the reports. However, it is evident that this type of knowledge use is a complex process that nonetheless has potential in that a clear majority of those questioned in this investigation feel that evaluations provide knowledge. The kind of “impact” that an evaluation could provide is difficult to assess. Investigations that are in close contact with organizations are necessary in order to further clarify the connection between the need for evaluations, relationships between the orderer and the evaluator, as well as the impressions of evaluations in knowledge use.

Discussion

Allow me to return to my initial line of questioning about the theory, methods and impacts of evaluation practice. The studied case “Metropolitan Initiative” has surely met the expectations of a case with high political relevance and a varied range of ways of working. The purpose with a case is not to create generalizable knowledge; we do not know whether the experiences from these evaluations can be transferred to other evaluation arenas. Nevertheless, it is possible to make some concluding observations. I will arrive at three observations in relation to theory, methods and influence.

Theory

A theory’s grade of selectivity, i.e. to what extent it is applicable and usable in different situations, is central to its survival (Payne 2005). Theories have different types of claims, therefore, the same type of problem can be understood in different ways. Their range is also limited: some theories claim to be linked to certain types of methods (for example evidence-based methods) and that they both explain and provide tools, while other theories only want to provide explanation.

Certain theories are adjustable and can be incorporated and combined to a more or less home-made constructed theory. How this is to be done, what parts that should be removed, and how they thus are put together to new theories can be important to try to clarify. Those theories that are utilized in connection with the evaluation in our case are often times not clearly expressed. On the contrary, independence toward the evaluation discourse is apparent which does not affect the evaluation to any greater degree. Then again it is seen that evaluations, albeit implicitly, should be conducted in close connection to the subjects of the evaluation and there is a general conviction that as an evaluator one should contribute to development and promote the evaluated activity. Perhaps this is the reason as to why evaluators many times are not clear with their scientific theoretical points of departure or with linking the work to fixed scientific positions and traditions of knowledge. The Metropolitan Initiative demonstrates a nearly pragmatic and functional approach toward evaluation theory. This is considerably more linked to social scientific and especially methodological points of departure than to that of specific evaluation theory which is hardly emphasized at all.

Methods

The Swedish evaluation tradition created in connection with the Metropolitan Initiative does not closely follow any of the models represented in the text books. Evaluators work without control and comparison groups and before/after analyses appear only occasionally and are far from a clear choice. Evaluators rather relate to public officials and public activities than to users and their organizations and prefer descriptive and process-focused studies.

Methodologically, there is in many cases a tendency to not put emphasis on explaining how work has been carried out – this is something that is implicitly understood as not being required in this form of study. A majority of studied evaluations do not live up to the methodological requirements placed on basic-level student work at a university. Instead, the typical evaluator looked for the legitimacy that can be found in dialogues and good relations with those who are being evaluated. Evaluators preferably work with non-structured interviews and with studying documents, but group interviews and observations are also represented. Outcomes can not be explained by traditional outcome measurements using control groups, but they are rather explained afterwards by trying thought processes and logical causal reasoning supported by qualitative data. Since the argumentation for the quality of these data normally is limited, the evaluations are dependent on the trust that evaluators have built up with their informants. This entails that the reliability of the evaluations is weak when they are assessed on the basis of scientific standards yet could be strong on the basis of local criteria.

Influence

It is clear, despite some apparent short-comings and frustrations from bad temporal timing, that the evaluations contribute to increased knowledge in those situations where public officials and politicians read them. It is apparent that they do this to a fairly high degree. On the other hand, surveys can hardly demonstrate to what degree evaluations make an impression within organizations and to what degree they are used as a basis for decision-making. More probing studies are necessary to be able to answer that.

The collected material nevertheless shows that an instrumental application occurs in nearly 30% of the evaluations and that even legitimating use occur. Currently, the evaluation constitutes an important and requested foundation when steering complex processes and it gives an opportunity for insight in the course of events which are otherwise difficult to grasp. Nothing in the material advocates that the attention surrounding the evaluation should decrease even if everyone is not on the same terms about how the knowledge should be utilized. On the contrary, the expectations are continuously high and the competence of the orderers in the seven municipalities has increased.

Conclusion: a narrative model for evaluation?

In conclusion, this investigation shows that evaluators apply what could be called a narrative model for evaluation. The evaluations are markedly qualitatively oriented and evaluators readily work in close connection to the projects that have been evaluated and they want to support the development of the projects. Further, theories and methods from the evaluation discourse are applied to a lesser degree and the reports suffer from obvious scientific quality problems. In the practical discourse the evaluating is appreciated by those subject to it. New knowledge is gained but it is unclear in what way the knowledge can be used. The orderer culture is so weakly developed that the orderers basically go by their instincts. Evaluators tell stories, something which in the big picture has been appreciated by the audience. There are always lessons to be learned. However, for a researcher whose task is to examine these stories from a scientific discourse the lack of quality is striking.

References

- Alkin, Marvin C (ed) 2004. *Evaluation Roots. Tracing theorists' views and influences*. Thousand Oaks: Sage
- Andersson, Roger 1997. *Förstudie avseende den nationella utvärderingen av statens insatser i invandrartäta bostadsområden*. Uppsala: Kulturgeografiska institutionen
- Argyris, Chris 1985. *Strategy, change and defensive routines*. London: Pitman.
- Blom, Björn & Stefan Morén 2003. *Insatser och resultat. Om utvärdering i socialt arbete*. Umeå: Institutionen för socialt arbete
- Brante, Thomas 1984. *Vetenskapens sociala grunder. En studie av konflikter i forskarvärlden*. Stockholm: Rabén & Sjögren.
- Bunar, Nihad 2004. "Det går bra men det ser fortsatt dåligt ut" *Organisationen och tillämpningen av lokala utvecklingsavtal i Stockholms kommun*. Huddinge: Södertörns Högskola
- Dahllöf, Urban 1989. Tredje gången gillt: en ny start för svensk utvärdering. I Sigbrit Franke-Wikberg (red) *Skolan och utvärderingen - fem professorer tar ordet*. Stockholm: HSL
- Davies, H., Nutley S. and Smith P. (eds) (2000) *What works? Evidence-based Policy and Practice in Public Services*, Bristol, Policy Press.
- Edström, Nina & Emil Plisch 2005. *En känsla av delaktighet. En studie av underifrånperspektivet i Storstadssatsningen*. Tumba: MKC
- Eliasson, Rosmari & Per Nygren 1980 *Utvärderingsforskning - kritisk vetenskap eller professionaliserad reformism*. Ur Armelius (red) *Ger behandling resultat?* Sthlm: Norstedts
- Eliasson, Rosmari & Claes Levin & Anna Meeuwisse & Sune Sunesson (red) 1990. *Den värderande blicken*. Lund: Studentlitteratur.
- Eriksson, Bengt & Karlsson, Per-Åke 1998. *Utvärdera bättre. För kvalitetsutveckling i socialtjänsten*. Stockholm: Gothia
- Fetterman, David 2001. *Foundations of Empowerment Evaluation*. Thousand Oaks: Sage
- Fishman, Daniel B 1999. *The Case for Pragmatic Psychology*. New York: New York University Press
- Franke-Wikberg, Sigbrit 1997. *Utvärderingens mångfald*. Stockholm: UHÄ
- Guba, E & Yvonna Lincoln 1989. *Fourth Generation Evaluation*. Thousand Oaks: Sage
- Healey, Patsy 1997. *Collaborative Planning. Shaping Places in Fragmented Cities*. London: Macmillan
- Henry, Gary T. & Melvin M. Mark (2003) Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions in *American Journal of Evaluation*, 24 (3), pp 293 – 312.
- House, Ernest R. & Kenneth R. Howe 1999. *Values in Evaluation and Social Research*. London: Sage
- Huw, Thomas & Patsy Healey (eds) 1991. *Dilemmas of Planning Practice. Ethics, Legitimacy and the Validation of Knowledge*. Aldershot: Avebury
- Jensen, Christian 2004. *Storstadssatsningens organisering*. Göteborg: Göteborgs universitet.
- Jerkedal, Åke 2001. *Utvärdering steg för steg. Om projekt- och programbedömning*. Stockholm: Norstedts Juridik
- Karlsson, Ove 1995. *Att utvärdera mot vad?* Stockholm: HSL Förlag
- Karlsson, Ove 1999. *Utvärdering mer än metod. En översikt*. Stockholm: Kommentus
- Kirkhart, Karen E. 2001. Reconceptualizing Evaluation Use: An Integrated Theory of Influence. In *New Directions for Evaluation*. Nr 88 2001
- Krogstrup, H (2003) *Evalueringsmodeller – Evaluering på det sociala område*. Aarhus: Systime
- Lindblom, Charles E. 1959. The science of "muddling through". *Public Administration Review* 1959. Ur Faludi, Andreas (ed) 1973. *A reader in planning theory*. Oxford: Pergamon Press.
- Nilsson, Kjell 1992. *Policy, interest and power. Studies in strategies of research utilization*. Lund: Meddelanden från socialhögskolan 1992:1
- Nilsson, Kjell & Sunesson, Sune 1988. *Konflikt, kontroll, expertis*. Lund: Arkiv.
- Nilstun, Tore & Göran Hermerén 1984. *Utvärderingsforskning och rättsliga reformer. Analys av orsaker och effekter*. Lund: Studentlitteratur.
- OECD 1998. *Integrating Distressed Urban Areas*.
- Patton, Michael Q (et al) 1977. In search of impact: An analysis of the utilization of federal health evaluation research. I Weiss, Carol (ed). *Using social research in public policy making*. Lexington.
- Payne, Malcolm (2000) *Modern Social Work Theory*. Houndmills: Palgrave.
- Premfors, Rune 1989. *Policyanalys*. Lund: Studentlitteratur.
- Qureshi H. (2004) 'Evidence in policy and practice. What kind of research designs?', *Journal of Social Work*, 4(1), pp. 7 – 23.

- Rombach, Björn & Kerstin Sahlin - Andersson 1995. *Från sanningsägande till styrmedel. Moderna utvärderingar i offentlig sektor*. Stockholm: Nerenius & Santérus förlag
- Sahlin-Andersson, Kerstin. 1989. *Okklarhetens strategi. Organiseringen av projektarbete*. Lund: Studentlitteratur.
- Scriven, Michael, 1991. *Evaluation Thesaurus*. Newbury Park: Sage
- Scriven, Michael 1998. The New Science of Evaluation, i *Scandinavian Journal of Social Welfare* 1998:2
- Shadish Jr, William & Thomas D. Cook & Laura C. Leviton 1991. *Foundations of Program Evaluation. Theories of Practice*. Newbury Park: Sage
- Simon, Herbert. A. 1976 3 ed (1945). *Administrative behavior. A study of decision-making processes in administrative organization*. New York: The Free Press.
- Socialstyrelsen 1995:19. *Att utveckla kvaliteten i socialtjänsten*.
- Socialstyrelsen 2000. *Nationellt stöd för kunskapsutveckling inom socialtjänsten..* Stockholm; Socialstyrelsen
- SOU 1998:25. *Tre städer. En storstadspolitik för hela landet*.
- SOU 2005:29. *Storstad i rörelse*.
- Soydan, Haluk 1998. Guest editorial, i *Scandinavian Journal of Social Welfare* 1998:2
- Stufflebeam, Daniel 2001. Evaluation Models, i *New Directions for Evaluation* 2001:89
- Tengvald, Karin 1995. *Kompetens och kunskapsutveckling*. Bilaga 7 i SOU 1995:59
- Trinder L. & Reynolds S. 2000. *Evidence-Based Practice. A Critical Appraisal*, Oxford, Blackwell Science.
- Triseliotes, John 1998. When is evaluation a scientific activity and when is it not? I *Scandinavian Journal of Social Welfare* 1998:2
- Vedung, Evert 1994. Utvärdering och de sex användningarna, i Rombach, Björn & Kerstin Sahlin – Andersson, *Från sanningsägande till styrmedel. Moderna utvärderingar i offentlig sektor*. Stockholm: Nerenius & Santérus förlag
- Vedung, Evert 1998. *Utvärdering i politik och förvaltning*. Lund: Studentlitteratur.
- Vedung, Evert 2002. Utvärderingsmodeller, i *Socialvetenskaplig Tidskrift*, vol 2-3, pp 118 – 143.
- Vedung, Evert & Jan-Eric Furubo & Rolf Sandahl 2000. Utvärdering i det svenska politiska systemet, i *Nordisk Administrativ Tidskrift* vol 2, pp 113 – 130.
- Wildavsky, Aaron 1979. *Speaking truth to power. The art and craft of policy analysis*. Boston: Little Brown and Company.
- Weiss, Carol 1980. Knowledge creep and decision accretion, i *Knowledge: Creation, diffusion, utilization*. Vol 1, no 3, March 1980 sid 381 - 404.
- Weiss, Carol 1998. *Evaluation*. Upper Saddle River: Prentice Hall
- Weiss, Carol & Bucuvalas, Michael 1977. The challenge of social research to decision making. I *Using social research in public policy making*. Weiss, Carol (red). Lexington.
- Yin, Robert K. 1989 (1984). *Case study research. Design and methods*. London: Sage.
- Åberg, Jan-Olof 1997. *Det rationella och det legitima. En studie av utvärderingars teori och praktik*. Göteborg: Sociologiska institutionen