Four Waves of Evaluation Diffusion
Evert Vedung
Evaluation 2010 16: 263
DOI: 10.1177/1356389010372452

The online version of this article can be found at:
http://evi.sagepub.com/content/16/3/263

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
The Tavistock Institute

Additional services and information for Evaluation can be found at:

Email Alerts: http://evi.sagepub.com/cgi/alerts
Subscriptions: http://evi.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav
Citations: http://evi.sagepub.com/content/16/3/263.refs.html

>> Version of Record - Jul 21, 2010

What is This?
Four Waves of Evaluation Diffusion

Evert Vedung
Uppsala University, Sweden

Abstract
This article investigates the dissemination of evaluation as it appears from a Swedish and to a lesser extent an Atlantic vantage point since 1960. Four waves have deposited sediments, which form present-day evaluative activities. The scientific wave entailed that academics should test, through two-group experimentation, appropriate means to reach externally set, admittedly subjective, goals. Public decision-makers were then supposed to roll out the most effective means. Faith in scientific evaluation eroded in the early 1970s. It has since been argued that evaluation should be participatory and non-experimental, with information being elicited from users, operators, managers and other stakeholders through discussions. In this way, the dialogue-oriented wave entered the scene. Then the neo-liberal wave from around 1980 pushed for market orientation. Deregulation, privatization, contracting-out, efficiency and customer influence became key phrases. Evaluation as accountability, value for money and customer satisfaction was recommended. Under the slogan ‘What matters is what works’ the evidence-based wave implies a renaissance for scientific experimentation.

Keywords
dialogue-oriented evaluation, diffusion of evaluation, evidence-based evaluation, neo-liberal evaluation, science-driven evaluation

Evaluation: A Popular Governance Recipe
Evaluation is an incredibly widespread governance formula. Since around 1990, the evaluation business has completely exploded. Virtually every intervention is a candidate for evaluation. The simple message is this. If you carefully examine and assess the results of what you have done and the paths toward them, you will be better able to orient forward. Good intentions, increased funding and exciting visions are not enough; it is real results that count. The public sector must deliver. It must produce value for money.

Evaluation as a governance strategy takes a myriad of forms. It manifests itself as goal-achievement evaluation, stakeholder evaluation, client-oriented evaluation, professional evaluation, self-assessment, randomized experimentation and quality assurance, to name but a few. Evaluation has become part

Corresponding author:
Evert Vedung, Uppsala University, Sweden.
Email: evert.vedung@ibf.uu.se
of larger management doctrines such as performance management, purchaser–provider models, partnership management and evidence-based management.

**Topic, Three Arguments and Limitations**

The topic of the present article is the major trends in the history of evaluation as they appear from a Swedish and to a lesser extent a North Atlantic vantage point from around 1960. Much as I realize that at least a few of these trends capture developments in countries like the Netherlands, the United Kingdom, the USA, Canada and to some extent Denmark and Finland as well, I shall not overstretch my observations in this context. Instead I hope that my article will inspire other, more knowledgeable scholars to contribute evaluation histories as well. First, I argue that the present evaluation landscape is shaped by four different waves of evaluation. Second, I maintain that these waves are parts of even larger waves, in which evaluation has been coupled to diverse, more general public sector governance doctrines. Third, I claim that behind these waves of evaluation-coupled-to-governance doctrines, there have been strong currents from both the left and the right of the political spectrum. There the article stops. Within the confines of a brief journal article, there is no space to delve deeper into the admittedly fascinating subject of the driving forces behind the various waves. While it is well known that evaluation has been received much earlier in some policy sectors than others, the article does not address that issue. Nor is the article an attempt to critique the contents of the various waves. The purpose is limited to outlining evaluation history from a Swedish and, to a much lesser extent, a North Atlantic vantage point (for my earlier efforts, see Vedung, 1992, 2000, 2004).

**Evaluation: A Very Simple Idea**

Evaluation rests on a very simple idea. Traditionally, public activities gained acceptance through proper procedures and strong economic investments, combined with beautiful rhetoric seasoned with reference to noble principles, the best of intentions and decent goals. Rarely was the public sector legitimized by reference to achieved results. Yet proponents of evaluation argue this is no longer sufficient. What count are actual achievements. Public policy must become results oriented. The evaluation megatrend revolves around systematic assessment and reporting of results as well as the implementation of ongoing or recently completed public activities in addition to, or at the expense of, on the one hand, control of whether rules of conduct are followed in the agencies and, on the other, future-oriented planning of pending operations. In some places, starting around 1965, there has been a noticeable shift of interest in policy communities from process-oriented management and analysis *ex ante* of probable consequences of proposed interventions to *ex post* analysis of already adopted or finished interventions.

**Evaluation: A Minimal Definition**

Evaluation is a nebulous concept. In this article, evaluation is minimally defined as careful retrospective assessment of public-sector interventions, their organization, content, implementation and outputs or outcomes, which is intended to play a role in future practical situations. In one sense, this definition is narrow. Evaluation is primarily concerned with interventions, i.e. actions of some kind that are taken to influence the world, and not with persons, commodities or states of the world. Evaluation is equivalent to *ex post* evaluation, i.e. assessment of adopted, ongoing or finished interventions. It excludes evaluation *ex ante*: calculated appraisals of the consequences of proposed and
considered interventions that are performed \textit{before} interventions are adopted and put into practice. It should be noted that assessments performed on interventions in empirical pilot trials are included in the evaluation category. Also, groupings of several evaluations into meta-evaluations are considered evaluations. In another sense, the definition is wide. It is not limited only to effects of interventions and activities at the outcome level (i.e. in society or nature) but also includes outputs, implementation processes, content and organization (Vedung, 1997: 2–13; 2006: 397).

**Four Evaluation Waves and their Depositions**

Anyone who has followed the field for more than 35 years will have noticed that several of the tenets of the evaluation movement have been more celebrated at certain times than at others. We have also witnessed that one form of evaluation has sometimes lost much of its prestige and support in favour of new forms. But these new forms have in turn lost their prestige and support in favour of other fads (Furubo and Sandahl, 2002).

The whole situation can be likened to ocean waves that roar in and subside, roar in and subside. We can therefore speak of evaluation waves that have swept ashore. In subsiding, the waves have not disappeared without trace but have left behind layers of sediment. In due time, the evaluation landscape has come to consist of layers upon layers of sediment.

I have discerned four evaluation waves with concomitant accumulations of sediment: the science-driven wave, the dialogue-oriented wave, the neo-liberal wave and the evidence wave. Each of these waves has formed part of a larger wave that also contains broad governance doctrines to which the evaluation forms have been coupled.

**The Science-Driven Wave**

From its inception in the late 1950s and consolidation in the mid-1960s, evaluation has been embedded into one of the great narratives of our time: that the world can be made more humane if capitalism and the market economy can be reined in by appropriate doses of central policy planning and public intervention at a comprehensive level. After 1965 this narrative was carried forward by a very strong wind from the political left, which manifested itself in the 1968 student revolts against the capitalist consumer society and the huge demonstrations against the US war on the socialist-oriented FNL guerrillas in Vietnam. The narrative included a strong admiration for Mao Zedong’s Chinese road to socialism and energetic demands for a socialist transformation of the capitalist market systems in the West.

In public-sector thinking, this was hailed as a victory of a kind of rationality. Public policy should be made more scientific and sensible (Hajer and Wagenaar, 2003, Radin, 2000). One name for this train of thought is \textit{radical rationalism}. Evaluation emerged as an element of radical rationalism.

Contemporary evaluation is in part a legacy of one element of this mighty current of substantive rationalism, which ran high in some western countries between approximately 1965 and 1975. To become more rational, public decision-making bodies should exploit the full arsenal of methods for programme budgeting, zero-based budgeting, multi-annual planning, futures studies, systems analysis and cost-benefit analysis, which are sometimes jointly called ‘policy analysis’ (for an excellent overview, see Premfors, 1989; Radin, 2000). Through the use of research and science-like analysis, sensible and long-term policies should significantly affect, if not completely replace, the short-term games between parties and interest groups, who by means of rules of thumb, passing fancies and anecdotal knowledge attempt to muddle through societal problems (Wittrock and Lindström, 1984).
In emphasizing the future-oriented, planning stages of public decision-making processes, radical rationalism was firmly oriented towards the front end of the policy cycle. The important thing was that particularly very large comprehensive public interventions were carefully designed in central planning machinery before they were adopted; adoption, enforcement and practical effectiveness were perceived as being relatively unproblematic (Wittrock and Lindström, 1984).

Radical rationalism stressed that problems should be examined in their entirety before governments decided to intervene. Major decisions should be taken only after scientific or science-like policy analyses had provided answers to a series of questions, which are summarized in Table 1.

During the 1960s and early 1970s, advocates of science-infused central planning celebrated triumphs all over the democratic world. ‘New planning systems emerged as mushrooms after rain’, to quote Ståhlberg (1986: 75).

The Engineering Model of Public Intervention

In the most extravagant form of radical rationalism, academic evaluation research would bring well-underpinned knowledge of actual effects of interventions to the relevant decision-makers for their consideration. The interaction between the decision-making and the scientific evaluation communities was fashioned according to the so-called engineering model.

The engineering model of public governance – perceived as an ideal to pursue – was a powerful driving force behind the emergence of the notion that evaluation may be a way of curbing the demons of political life to reduce its irrationality (Albæk, 1988: 23ff.). In any case, it played a role in various semi-technocratic conceptions that government decision-making must be guided by feedback from scientific evaluation in order to be demystified and become more rational. The engineering model for use in public-sector evaluation is reproduced in Figure 1.

A problem in society or nature is discovered and framed, for example by researchers, mass media or interest groups. The extent of the problem, its causes and consequences are analysed. The problem with its causes and consequences is put on the political agenda. Then the decision-makers set goals (ends, targets) for what they want to achieve in the problem field. Next, they decide to launch a provisional small-scale tryout. This tryout is contracted out to evaluation researchers in academia. Acting as distanced neutral observers and armed with the best available scientific method in the form of randomized (or matched) two-group experimentation, the researchers empirically test what means (instruments and other efforts) are most efficient (or effective) to achieve the

### Table 1. Steps in Radical Rationalism

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>What problem is the decision-maker seeking to resolve?</td>
</tr>
<tr>
<td>2.</td>
<td>What are the causes of, or the driving forces behind, the problem?</td>
</tr>
<tr>
<td>3.</td>
<td>What impacts is the problem likely to have, if nothing is done?</td>
</tr>
<tr>
<td>4.</td>
<td>What goals have been set in the problem area?</td>
</tr>
<tr>
<td>5.</td>
<td>What alternative means are thought to contribute to the attainment of the goals?</td>
</tr>
<tr>
<td>6.</td>
<td>What are the consequences of the alternative means and what is the probability of these consequences being generated?</td>
</tr>
<tr>
<td>7.</td>
<td>What costs and resource requirements are associated with the different means?</td>
</tr>
<tr>
<td>8.</td>
<td>In light of these calculations concerning consequences and costs, how can the alternative means be arranged, and what criterion should be used in the selection of means?</td>
</tr>
<tr>
<td>9.</td>
<td>How can experiences from choices already made and implemented be used to reassess the goals set and the means adopted?</td>
</tr>
</tbody>
</table>

(Extended and modified from Wittrock and Lindström 1984)
goals set. The science-based findings from these trials are fed back to the decision-makers, who make a proper binding decision to impose the most efficient of the examined interventions in full scale (instrumental use). The intervention decision is then submitted to managers and operators, who neutrally and faithfully implement it to produce the desired outcome.

According to the engineering model, intervention decisions should be taken in two quite clearly discernible stages. The first, preliminary stage suggests that conceivable measures to reach given ends should be rigorously tested in carefully designed, small-scale pilot trials. The findings of the pilot trials should be fed back into the political system, which in a second stage, on the basis of the findings, should arrive at a decision about the full-scale introduction of the most effective measure to achieve the stated ends.

The engineering model posits that evaluative findings are used instrumentally. By *instrumental use* is meant that evaluation discoveries about means are accepted as true and transformed into binding decisions. Evaluation is neutral, objective research. It does not formulate problems. Nor does it recommend ends (goals). The function of evaluation before an intervention is introduced on a full scale is to help determine the most efficient means of achieving the previously stated ends, that is, the means that will ensure goal achievement at the lowest cost. In the engineering model, knowledge of means is produced in a value-neutral fashion, given that the evaluation meets the highest scientific standards. For the ends have been set by the decision-makers, and finding the most efficient means to reach these externally set ends (goals) is regarded as a proper task for objective, empirical research (Simon, 1976: 37).
During the 1960s and even earlier, advanced evaluative thinking and practice was driven by this notion of scientification of public policy and public administration. Evaluation would make government more rational, scientific and grounded in facts. Evaluation was to be performed by professional academic researchers.

The Dialogue-Oriented Wave

Towards the mid-1970s, confidence in experimental evaluation faded. Evaluation should be more pluralistic, it was argued. Participants other than politicians, upper management and academic researchers should be involved. All stakeholders in an intervention should be activated (Guba and Lincoln, 1989; Karlsson, 1995).

Although much older than evaluation, the stakeholder idea was incorporated into the evaluation discourse and practice at about this time. Stakeholders were defined as groups or individual actors that have some interest vested in the intervention to be evaluated. Interest may be measured in terms of money, status, power, face, opportunity or other coin, and may be large or small, as constructed by the groups in question (Guba and Lincoln, 1989: 51). A list of potential stakeholders for, as an example, a European local public-sector hospital, is shown in Table 2.

Assessments should be set up as stakeholder evaluations with significant stakeholding audiences represented. Evaluation should continue to be a concern for politicians and top-level managers but now as members of a larger group of stakeholders communicating with the evaluators and each other and including operators like doctors, nurses and non-medical employees, target group members and relatives of target group members. The claims, concerns and issues of the various stakeholders should serve as points of departure for evaluations. Far from being carried out as rigorous scientific two-group experimentation, stakeholder evaluation was supposed to be conducted by discussion, dialogue and communication among equals, even deliberation avant la lettre (see e.g. Guba and Lincoln, 1989: 56–7). Thus, the dialogue-oriented wave is an appropriate designation. The proponents themselves often called it ‘democratic evaluation’.

Yet, the process among stakeholders should engender more than just dialogue and communication. Guba and Lincoln (1989: 56–7) put it in this way:

The involvement of stakeholders … implies more than simply identifying them and finding out what their claims, concerns and issues are. Each group is required to confront and take account of the inputs from other groups. It is not mandated that they accept the opinions and judgments of others, of course, but it is required

<table>
<thead>
<tr>
<th>Table 2. Stakeholders in a European Local Public Sector Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Patients (=clients, users, target group members)</td>
</tr>
<tr>
<td>• Patients’ relatives</td>
</tr>
<tr>
<td>• Doctors, nurses</td>
</tr>
<tr>
<td>• Non-medical employees</td>
</tr>
<tr>
<td>• Hospital upper management</td>
</tr>
<tr>
<td>• Suppliers of medical equipment</td>
</tr>
<tr>
<td>• Regional Parliament (politicians)</td>
</tr>
<tr>
<td>• Regional health administration upper management</td>
</tr>
<tr>
<td>• Municipal council (politicians)</td>
</tr>
<tr>
<td>• Municipal social commission</td>
</tr>
<tr>
<td>• Municipal social services</td>
</tr>
</tbody>
</table>
that they deal with points of difference or conflict, either reconstructing their own constructions sufficiently to accommodate the differences or devising meaningful arguments for why the others’ propositions should not be entertained.

In this process a great deal of learning takes place. On the one hand, each stakeholder group comes to understand its own construction better, and to revise it in ways that make it more informed and sophisticated than it was prior to the evaluation experience … On the other hand, each stakeholder group comes to understand the constructions of other groups better than before. Again we stress that that does not mean coming to agreement, but it does mean gaining superior knowledge of the elements included in others’ constructions and superior understanding of the rationale for their inclusion.

… [a]ll parties can be mutually educated to more informed and sophisticated personal constructions as well as an enhanced appreciation of the constructions of others.

Actually, for this wave of evaluation, Guba and Lincoln (1989: 43ff., 83ff.) proposed an alternative, constructivist paradigm to the conventional, positivist, scientific paradigm in which the science-driven wave was grounded (see also Dahler-Larsen, 2001). ‘It rests in a belief system that is virtually opposite to that of science’, the authors argued.

Also labelled the naturalistic, hermeneutic or interpretive paradigm, the constructivist paradigm was different from the positivist paradigm at three levels: ontology, epistemology and methodology. As regards **ontology** (what is the nature of reality?) the constructivist paradigm denies the existence of objective reality, asserting instead that realities are social constructions of the mind and that there exist as many such constructions as there are individuals. There is no objective truth on which inquiries can converge.

As regards **epistemology** (how can we be sure that we know what we know?), the constructivist paradigm denies the possibility of subject–object dualism, suggesting instead that the findings of a study exist because there is an interaction between observer and observed that literally creates what emerges from that inquiry. One cannot find out the truth about how things really are or how they really work. It is impossible to separate the inquirer from the one being inquired into. It is precisely their interaction that creates the data that will emerge from the inquiry.

As regards **methodology** (what are the ways of finding out knowledge?), the constructivist paradigm cannot use approximation to reality as the criterion to ascertain which construction is better than others because the possibility of an objective reality is denied. Instead, it uses a hermeneutic-dialectic process. As Guba and Lincoln (1989: 89–90) argue:

> [A] process must be instituted that first iterates the variety of constructions (the sense-making) that already exist, then analyzes those constructions to make their elements plain and communicable to others, solicits critiques for each construction from the holders of others, reiterates the constructions in light of new information or new levels of sophistication that may have been introduced, reanalyzes, and so on to consensus – or as close to consensus as one can manage. The process is hermeneutic in that it is aimed toward developing improved (joint) constructions … It is dialectic in that it involves the juxtaposition of conflicting ideas, forcing reconsideration of previous positions.

The science-driven wave rested upon **means-ends rationality**, emanating from the thinking of Max Weber. Given that goals and objectives were set by bodies outside the scientific community and expressly recognized as subjective, academic research could examine in experimental settings the ability of various means to reach these externally set ends. Experiments would deliver objective
generalized truths about means. Other names for this train of thought in the social sciences are behaviouralism and positivism. In contrast to the science-driven wave, the dialogue-oriented wave rested upon communicative rationality. Instead of producing truths, dialogical evaluation would generate broad agreements, consensus, political acceptability and democratic legitimacy.

Already during the final years of the 1960s, new social movements emerged that criticized elitist central societal planning. Coming from the left of centre in the political spectrum, the criticism also contained environmentalist tenets. Over time, the dialogue-oriented wave developed towards a participatory criticism of extant representative democratic government.

In the early 1990s, proponents of the dialogical wave started to raise demands for more ‘deliberative democracy’. Representative democracy, with its general elections and concomitant expert policy analyses should be supplemented by venues for serious policy communication among ordinary people. Election campaigns and so-called ‘debates’ in municipal councils and other parliamentary arenas mostly declined into ‘pie-throwing’, it was argued. New forums were needed, deep down in systems where clients and other stakeholders could meet for serious discussions of existing public interventions and proposals for new action. Evaluations were presented as appropriate arenas for deliberative democracy, which might deepen representative democracy. The point was that users and other stakeholders, by participating in such evaluative dialogues, as a long-term side-effect, would learn to become more engaged and better citizens, and thereby strengthen representative democracy (Sjöblom, 2003). The trend towards ‘empowerment’ and ‘empowerment evaluation’ can be traced to this time, too.

The Neo-Liberal Wave

One might put the NPM ideals very simply as a desire to replace the presumed inefficiency of hierarchical bureaucracy with the presumed efficiency of markets. (Power, 1994: 43)

Rarely does a turn of events fulfil its concomitant expectations. The pioneers of evaluation in the 1960s claimed that the problem with the system of representative democracy and the public sector was that it was based too much on biased ideological beliefs, political tactics, pointless bickering, passing fancies and anecdotal knowledge. The cure was a strong dose of science focusing specifically on intervention effects. Unbiased evaluation research would eradicate the aberrations of the representative system of democratic governance. Yet, as we have seen, this science-inspired movement soon ran into demands for more stakeholder involvement and dialogue and communication among concerned stakeholders, interest groups and citizens. Both of these tendencies drew their strength from the left wing of the political spectrum.

Around 1978–9, a third swing started to sway the field of evaluation. Politically, the new Zeitgeist implied a turn to the right. Its banner was neo-liberal, its contents were confidence in customer orientation and markets. What was novel was not that goal achievement, effectiveness, efficiency and productivity became catch phrases but that these objectives were to be achieved by government marketization instead of stakeholder involvement or scientification from the top down. Decentralization, deregulation, privatization, civil society and in particular, customer orientation became new slogans. Previously regarded as the solution to problems, the public sector now became the problem to be resolved.

The collective term for the neo-liberal public sector reform movement is New Public Management. (Hood, 1991; Hood and Jackson, 1991; Klausen and Ståhlberg, 1998; Osborne and Gaebler, 1992; Pollitt, 2003; Pollitt and Bouckaert, 2004; Pollitt et al., 1999). More focus on results, less focus on processes, was the fundamental idea in New Public Management. Under this
umbrella, New Public Management harboured a cluster of ideas drawn from administrative practices in the private sector. The main dogmas of New Public Management are shown in Figure 2.

New Public Management contains three major elements. The first element is belief in leadership. ‘Let managers manage’ is the battle cry. It is with leadership at the centre that new and more dynamic results-oriented organizations should be created. The margin for leadership should increase. This applies to political as well as agency leadership; to leadership in educational as well as in service institutions. Leadership should be exercised by management professionals. Good leadership must be taught and learned. Being an expert on the actual substantive issues is not enough. Leaders must meet demands for performance, efficiency and other management skills.

The second element involves increased use of indirect instead of direct control. Total privatization is included, but only as one tenet (Osborne and Gaebler, 1992: 45: ‘privatization is one arrow in the government’s quiver’). A major point is that the government should act as the helmsman of the ship of state, but not necessarily as an oarsman (‘steering not rowing’, Osborne and Gaebler, 1992: 25). In any case, the steering and rowing functions should be separated. Corporatization and outsourcing of public services as well as increased competition are also important to boost flexibility, avoid wastage and counterbalance public employees’ self-interest.

The best-known feature of New Public Management is results-based management (ESV, 1999: 20; Kusek and Rist, 2004; Perrin, 1998; Pihlgren and Svensson, 1989; Sandahl, 1992). In its pure ideal-type form, results-based management (performance-based management, management by objectives) is a process of several steps (Table 3).
The executive self-selection of means (item 5 in Table 3) is an indispensable tenet of results-based management. A second essential element is the central position accorded to monitoring and follow-up. Feedback of information on the extent to which the objectives of the various outcome stages are met is required from the agent and thought to play a key role once responsible principals at different levels assess where the whole system stands. These data also form the basis of an effects-oriented process and outcome evaluation, which is a third building block. The findings of the monitoring and evaluation efforts are then to be used to highlight successful authorities and through the power of the example, spur less successful authorities to increased efforts – a fourth ingredient. This strong shift towards ex-post assessment and the back-end of the intervention cycle is typical of administrative thinking from around 1980.

During the pioneer era in the 1960s, evaluation was seen as a one-off enterprise, resorted to in times of urgent need only. New Public Management aspired to change this and turn evaluation into a permanent feature of the larger doctrine of results-based management.

An important trend in this context is outsourcing (contracting-out). After a phase of competition for tenders among various private and public providers, public agencies commission some outside body to carry out certain tasks for which the public sector is responsible. Examples are distribution of food, provision of hot water, computer support and child care. The public sector should be a catalyst, a starter, a broker, but not necessarily a principal service provider. Since contracts are written on a regular basis, the public sector may gain a flexibility they might not have if they produced everything. This means that new purchaser–provider relations are created. New agents (private and public) will enter the scene, which the principal (government or other public authorities) may need to supervise. This supervision may be a task for evaluation. Evaluations in these cases are primarily tools for accountability, not for promotion.

The third NPM element is customer focus. This part of New Public Management focuses on how organizations can be reformed so that actual and potential users and clients get more influence and performance becomes more customized. The idea is that the public system puts too little emphasis on the preferences, needs and interests of the citizens and users. When there is no market

---

**Table 3. Steps in Results-Based Management**

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The principal, for example the superior national authority, creates an overall vision and sets some clear outcome goals that indicate successive stages towards the overall vision.</td>
</tr>
<tr>
<td>2</td>
<td>The principal and the implementing agent, e.g. the municipality, together develop indicators of the successive outcome goals.</td>
</tr>
<tr>
<td>3</td>
<td>The agent develops indicators of measures that the target audiences, such as households, shall take and of its own final outputs.</td>
</tr>
<tr>
<td>4</td>
<td>The agent is awarded financial resources as an unspecified lump sum.</td>
</tr>
<tr>
<td>5</td>
<td>Within the budgetary frames set by the principal, the agent independently chooses the means (outputs) to achieve the goals.</td>
</tr>
<tr>
<td>6</td>
<td>The principal advertises that the indicators will be monitored and that evaluation in the form of effects evaluation will be carried out later.</td>
</tr>
<tr>
<td>7</td>
<td>To follow up, data on the indicators will be gathered, preferably by the agent who will also account for them to the principal.</td>
</tr>
<tr>
<td>8</td>
<td>The principal evaluates whether the goals and objectives have been attained and whether the means chosen by the agents have contributed to goal achievement; this may be done on the basis of data emerging from the follow-up.</td>
</tr>
<tr>
<td>9</td>
<td>The principal and the agent use the follow-up and the outcome-effects analysis (evaluation) to correct goals and means.</td>
</tr>
</tbody>
</table>
to signal demand for public services, other mechanisms are needed to improve the flow of information from the users of the intervention into decision-making processes. This can be achieved if users are able to choose between alternative service providers or participate on institution and agency boards. It can also be achieved through hearings and questionnaires. New Public Management stresses that the authorities have to be more responsive and adapt to customers. An evaluation moment is introduced when customer satisfaction with the service is measured, for example, with the help of customer satisfaction indices.

It should be stressed that NPM, in contrast to the dialogue-oriented wave, is customer oriented, not stakeholder oriented.

All this has not led to the disappearance of evaluation. In the neo-liberal wave, it is regarded as imperative that the fundamental principal in a representative democracy, the demos, has a right to know how her agents spend her money. This results in an increased emphasis on the accountability of agents in terms of resource use, by checking for economy, effectiveness and cost efficiency. Evaluation has thus been strengthened and, above all, taken on new forms. Evaluation has become a permanent feature of results-based management and of outsourcing. Evaluation has taken on new expressions in the form of accountability assessments, performance measurement and consumer satisfaction appraisal. Quality assurance and benchmarking are also recommended.

The Evidence Wave: The Return of Experimentation

Around 1995, a fourth evaluation wave started to roll in over the North Atlantic world and from the year 2000, the Nordic countries as well: the evidence movement. Supporters of this movement demand that government activities be based on success: ‘What matters is what works’. And what works is called evidence. Although it is growing in popularity, this fourth evaluation wave still is not as strong as the science-driven wave in the 1960s and the neo-liberal wave in the 1980s and 1990s.

In social work, public health, education, crime prevention, biodiversity and related fields, international cooperation bodies began to be established around 1995 to produce ‘systematic reviews’ of the evidence-based lessons learned. The lessons learned are about intervention consequences at outcome levels. Evidence is ranked on the basis of an evidence hierarchy. In such hierarchies, evaluative designs are graded according to their ability to causally produce safe knowledge of intervention effects. The example in Figure 3 is based on the work of Ray Pawson (2006: 49) and Rieper and Foss Hansen (2007).

Randomized controlled trials (with concealed allocation)
-------------
Quasi-experimental studies (using matching)
-------------
Before-and-after comparison
Cross-sectional, random sample studies
Process evaluation, formative studies and action research
Qualitative case studies and ethnographic research
Descriptive guides and examples of good practice
Professional and expert opinion
User opinion

Figure 3. Evidence Hierarchy in Meta-analysis
It is noteworthy that randomized experiments and quasi-experiments are ranked the highest, while user opinion of effects is ranked the lowest. This is a far cry from the strong client-orientation of New Public Management. From the English wordplay ‘evidence-based versus eminence-based medicine’ (*Times Literary Supplement*, 8 Feb. 2008) it is obvious that the evidence movement wants to play down professional judgements in favour of scientific experimentation (cf. Rieper and Foss Hansen’s excellent report (2007)). The evidence wave tends to structure the field from a social science methodology point of view, not a political, administrative or client-oriented one. Tacitly at least, it is based on *means–ends rationality*, where the task of evaluation is to enhance and disseminate knowledge of means. The evidence movement, some pundits argue, involves the return of science-based evaluation, but in a new disguise.

Most famous among the international cooperation bodies is the Campbell Collaboration, named after Donald T. Campbell, the celebrated advocate of a science-based, experimental public policy. Another global institution that works in the same spirit is the Cochrane Collaboration.

Typical for the operation of these new international bodies is that they do not carry out evaluations by themselves. Instead, they engage in meta-analysis, or synthesis analysis. The preferred term, mentioned above, is ‘systematic reviews’. The Centre for Evidence-based Conservation in Britain gives the following explication of the term ‘systematic review’:

Systematic review is a tool used to summarise, appraise and communicate the results and implications of a large quantity of research and information. It is particularly valuable as it can be used to synthesise results of many separate studies examining the same question, which may have conflicting findings. Meta-analysis is a statistical technique that may be used to integrate and summarise the results from individual studies within the systematic review, to generate a single summary estimate for the effect of an intervention on a subject.

The purpose of a systematic review is to provide the best available evidence on the likely outcomes of various actions and, if the evidence is unavailable, to highlight areas where further original research is required. It is, therefore, a tool to support decision-making by providing independent, unbiased and objective assessment of evidence.

The roles recommended for evaluators are also different from those of the science-driven wave in the 1960s. In the 1960s, academic evaluators would carry out experimentally designed evaluations commissioned by governmental bodies. In the evidence movement, two roles are open. One is that evaluation is conducted at universities as uncommissioned basic research. The second is that public-sector practitioners should conduct research on their clients as part of their clinical work. The ideal is the doctor who acts as a clinical practitioner towards her patients while also doing scientific research on them.

Can studies carried out in academia as basic research really be characterized as evaluations? The answer is yes. They revolve around interventions, i.e. they are action-oriented. They are carefully carried out, concentrated on intervention effects and intended for potential adoption as policies or programmes in public governance. And syntheses of several such studies are included in my minimal definition of evaluation at the outset of this article.

In the 1950s and 1960s, the idea that decisions on public policies and programmes should rest on scientific information came from the defence sector. The RAND Corporation in the US and the Defence Research Institute in Sweden were the initial proponents of this notion. This time, the impetus was provided by the medical field. It started with demands for evidence-based social medicine, which later led to cries for evidence-based social work, evidence-based public health and evidence-based crime prevention (Sherman, 2002).
Conclusions

Evaluation is currently an exceptionally fashionable management recipe. Virtually every public sector intervention is and should be evaluated.

Trees, Generations and Waves

In examining the development of evaluation, others have used the tree metaphor and the generation metaphor. As organic metaphors, both trees and generations may depict kinship. They also illustrate that several evaluation philosophies may coexist. Marvin C. Alkin has used the tree metaphor to illustrate the development of evaluation in the US (2004: 13). The generation metaphor captures the time dimension even better. Like generations of humans, several evaluation philosophies may exist at the same time. Guba and Lincoln famously divided evaluation development into four generations (1989). The wave metaphor may also capture the passage of time. I have used waves with their concomitant depositions of sediment to express the fact that several waves have swept over parts of the world at different times.

Four Evaluation Waves Coupled to Diverse Governance Doctrines

Four evaluation waves have swept across Sweden – and at least to some extent also other countries of the North Atlantic world – since around 1960: the scientific wave, the dialogue-oriented wave, the neo-liberal wave and the evidence wave. All four have deposited layer upon layer of sediments that have remained even when the next waves have rolled in (cf. Wollmann, 2003). Together these sediments have shaped current thinking and practice of evaluation.

Starting in the 1950s and becoming established in the middle of the 1960s, evaluation was part of a much larger stream of ideas to make government more scientific. The public sector would perform much better with a proper dose of trustworthy scientific findings about the real results of adopted policies and programmes, it was maintained. Given externally set goals, professional academic researchers should be commissioned to scientifically evaluate appropriate means to reach these goals through controlled two-group experimentation. Evaluation was based on means–ends rationality. While goals were considered subjective, the means to reach the goals could be ascertained in an objective, scientific way. The findings of the evaluations would then inform public decision-making.

The driving force behind this *scientific wave* was a broad and strong left-wing ideological current. The idea was to refine central planning and welfare state measures to eradicate poverty and make the mixed economy more effective and efficient.

Already in the early 1970s faith in methods-driven scientification of government started to languish. Mistrust of experimental evaluation gained momentum. Demands were voiced for more participation by diverse groups and more dialogue and communication in evaluations. Supporters of this *dialogue-oriented wave* often characterized it as democratic evaluation. *Representative democracy* might be supplemented by evaluative arenas where users, citizens and other stakeholding audiences deliberated effects and implementation of public interventions.

Although considerably older than evaluation, the stakeholder-dialogue idea was incorporated into evaluation discourse and practice at about this time. And it has stayed there ever since. This second wave was also driven by an ideological current from the left, but sprinkled with some green, environmentalist tenets.

Around 1980, the *neo-liberal wave* came rolling in, this time from the right of centre on the political scale. The neo-liberal wave was based on a mistrust of central planning but saw the
remedy not in dialogue and participation but in more market orientation. Deregulation, privatization, efficiency and customer orientation became new key words. Evaluation came to be included in a neo-liberal, market-oriented train of thought called New Public Management.

New Public Management pushed strongly for evaluation as accountability and value for money. Accountability evaluation became a permanent feature of performance management and outsourcing. Evaluation took on new expressions in the form of customer-oriented evaluation. Value-for-money evaluation in the form of cost-effectiveness and productivity studies was highly regarded.

While gaining in strength, the fourth evaluation wave is not yet as strong as the scientific wave of the 1960s and the neo-liberal wave that grew in popularity from 1980 onwards. Characteristic of this evidence wave is an effort to make government more scientific and based on real empirical evidence. It is concerned with what works. This can be interpreted as a renaissance of science and randomized experimentation. It is basically driven from the right-of-centre end of the political spectrum.

References

**Evert Vedung** is emeritus professor of political science, especially housing policy, at Uppsala University’s Institute for Housing and Urban Research and Department of Government. His works on evaluation in English include *Public Policy and Program Evaluation* (author, 1997 2000) and *Carrots, Sticks and Sermons* (1998, 2003, coeditor). Please address correspondence to: Uppsala University, IBF (Institute for Housing and Urban Research), PO Box 785, SE-801 29 Gävle, Sweden. [email: evert.vedung@ibf.uu.se]