

PRELIMINARY DRAFT

## ***Public Management and Evaluation***

*paper prepared for workshop 4 (Public Management)  
2005 NIG Annual Work Conference Nijmegen, November 11, 2005*

**by Frans-Bauke van der Meer<sup>1</sup>**

### ***Abstract***

In public management we can observe a continuous search for its own improvement. New arrangements and procedures, often derived from NPM ideas, but also from other PA insights, are put in place with the expectations that public administration as well as its related agencies and networks will function more efficient, effective, transparent, responsive or produce better quality afterwards. A key question, therefore, is whether and when PM reforms and innovations 'work'. Various forms of monitoring and evaluation can help to answer this question. As a matter of fact many public management reforms explicitly include assessment procedures and related feedback mechanisms. However, the PM reforms and their assessment mechanisms as such are evaluated far less frequent. This paper aims to explain why this is so and to suggest ways in which such (meta)evaluation can be realized. In doing so, two distinct 'logics' of evaluation are identified. The first logic essentially is the rationality of feedback mechanisms (systems theory) that is part of the 'repertoire' of much NPM thinking. The other logic is based on social scientific insights in the conditions for and mechanisms of utilization and impact of evaluation processes and results.

---

<sup>1</sup> dr. F.B. van der Meer is associate professor of Public Administration, Faculty of Social Sciences, Erasmus University Rotterdam. P.O. Box 1738, 3000 DR Rotterdam. Phone: +31 10 4082137; e-mail: [f.b.vandermeer@fsw.eur.nl](mailto:f.b.vandermeer@fsw.eur.nl)

## Introduction

In the last four decades or so, governments and governmental bodies in 'western' countries have come under increasing pressure, both societal and financial in nature. The quest for efficiency, efficacy, responsiveness, flexibility and accountability of government is a dominant one. It has given rise to numerous innovations both in public organization and in policymaking and implementation. These innovations aimed at improving government in some respects in order to accommodate 'external' and 'internal' pressures.

Not all innovations are considered successful. Some are replaced by others within a few years time, in which case a change of direction (e.g. on a centralization – decentralization dimension) is not uncommon. In other cases innovations are supplemented with other ones, e.g. autonomization followed by a growing intensity of central control. On the other hand there are successes, at least in the eyes of some actors involved. Such 'best practices', however, may be copied elsewhere with mixed results. Moreover authorities and agencies are frequently criticized (e.g. in cases of a disaster) for procedures and practice they have taken for granted for long. It may be concluded that – although many procedures and innovations are considered 'necessary' and 'unavoidable' – not enough is known about their working and impact.

In this paper I will discuss what evaluation does and can contribute to gain more insight in the dynamics and effects of public management strategies and reforms. To do so, I pose three key questions:

1. How is evaluation actually used in public management?
2. How can public management innovations be sensibly evaluated?
3. When and how may such evaluations have impact on public management in practice?

Of course, the first question can only be answered in a sketchy mode within the framework of this paper. However some observations are to be made to enable an answer to the second and third questions. In the next sections I will identify a number of typical public management innovations and discuss their underlying rationales, which will be related to the dominant use of evaluation in the context of each specific type of management innovation. Next I propose a theoretical framework to develop notions with respect to the second question. This theoretical framework also helps to give some answers to the third question.

## Innovations in public management

We can distinguish a number of different types of public management reforms, although there are interesting interconnections between them. We will turn to these 'hybrids' later on.

A first type, which is probably most widespread (Kickert 1997; Pollitt and Bouckaert 2000), comprises the utilization of private sector management techniques and organizational forms in the public sector. These innovations, often taken together under the heading of New Public Management (NPM), seek to improve efficiency and quality of public administration and service delivery by decentral autonomy and output control, frequently regulated by management contracts. The underlying rationale seems to be that an incentive structure that rewards efficient production and measurable output enhances motivation and creativity of individuals and groups and can replace much efficiency reducing hierarchical control. Moreover the decentralized nature of the new arrangements is thought to enhance flexibility, responsiveness and tailor-made solutions: a demand orientation, which is presupposed to be the main guide of private sector management. Taken together NPM reforms are expected to trigger a better use of public sector resources.

The second type focuses not so much on government or public sector production per se but on joint production of policies, public works and services with private sector enterprises and/or societal organizations (Teisman and Verheij 1996; Tops and Weterings 1998). The aims are in a sense similar to those of the first type: improving efficiency and quality of public production, although here public agencies are not the sole producers. Here too, there is a parallel with private sector strategies, where organizations initiate joint enterprises to realize

projects, develop new technology or penetrate new markets. The main rationale, however, is quite different from the first type. The problem is not that the public sector uses its resources in a sub-optimal way, but that government lacks key knowledge, financial and other resources to realize many projects and solutions to societal problems. Government is only one among many actors in a societal network of interdependencies. Only by cooperation effective results can be hoped for.

A third set of new strategies in public management seeks to involve citizens and societal organizations having an interest in certain policy issues or domains. This participation (Edelenbos 2000) may include the phase of problem definition, generation of possible solutions, a role of citizens and societal organizations in policy implementation and evaluation, and a shift in responsibility towards the citizens or private initiative. By definition there is some overlap with the second type of public management innovations since here too, non-governmental actors are involved in the production of policy and policy outcomes. However, there are other rationales for such participative strategies. One of them starts from the notion that societal problems often cannot be identified in an objective and unambiguous way. The public task, then, is not so much to solve 'the' problem, but to search for ways to accommodate different problem definitions at the same time. By involving individuals and groups having an interest in at least some of these definitions, the searching process is part of its solution. In and by interaction with other actors (among which governmental ones) actors may change or broaden their problem definitions and/or come to conceive of new solutions. Thus, eventual outcomes may in part be created by and gain the support of stakeholders. The role of government is mainly a facilitating one in producing solutions and support.

In this short and crude description of public management innovations, I mentioned some overlap in goals and practices between the types discerned. Of course, in practice, the rationales are not isolated from each other either. For example, in the new Dutch law on care ('Wet Maatschappelijke Ondersteuning') all three rationales and the traditional hierarchical way of thinking can be found:

- hierarchy: the municipalities are made responsible by law for societal care
- co-production: care should be delivered as much as possible by private organizations
- participation: care clients should participate in formulating care policies; clients should take responsibility for organizing their own care as much as possible
- new public management: care delivering organizations are contracted and evaluated according to agreed performance measures

Thus, it is not only worthwhile to ask how successful individual public management innovations are and how valid their rationales are but it is also interesting to see how different types of reforms influence each other and how their rationales interact in practice.

## **Evaluation in public management**

In the framework of NPM-like innovations evaluations play an important role. Assessment of performance or output and comparing it with initial targets or contractual commitments is central to the philosophy of this type of approach. Moreover, the felt need for public accountability of (autonomized) agencies has given rise to numerous review procedures and control mechanisms, which also involve the collection and judgment of data on performance and results (see e.g. Sanderson 2001, who describes local practices in the UK). Such practices contribute to an 'evaluative state' (Sanderson 2001: 303) in which performance measurement and accountability constitute the dominant perspective. With respect to the impact of these evaluations four observations can be made.

First, they influence and focus actual behavior. Since performance indicators generally are known before the actual performance takes place, they can and will guide this performance. If the performance of academic work is assessed by counting the number of publications in international refereed journals, the number of such publications and the time devoted to producing them will rise. In a sense, the impact precedes the actual evaluation. But there

are also ex post impacts as for example the allocation of new budgets is based on evaluation results.

Second, this type of evaluations may give rise to unintended or perverse effects. This is a consequence of the fact that performance indicators used generally are only partial operationalizations or proxies of the goals aimed at. This phenomenon is known as the performance paradox (Meyer and Gupta 1994).

Third, frequently, especially in the relation between 'principal' and 'agent' the amount and the specificity of data produced by agents are so large that their impact is greatly reduced by the overload it creates for the principal.

Finally, the fact that evaluations in this context focus on accountability may produce 'defensive' reactions in the sense of denying the validity of evaluation outcomes or seeking excuses beyond control. Such reactions reduce the opportunities for learning since the question of possible improvements is evaded.

Now, a relevant question is what the implications of the potentially negative effects of evaluation in NPM-contexts are. Can evaluation procedures be designed in a way that prevents such undesired consequences? Or are we dealing here with a fundamental shortcoming in the rationale behind NPM? In view of the impressive amount of evaluations produced in relation to NPM-like management, it is remarkable that NPM-reforms as such, including the evaluation procedures involved, are evaluated only incidentally (Pollitt 1995; Pollitt and Bouckaert 2003; Forbes and Jr. 2005). In a later section I will propose an explanation for this feature.

In connection to co-production and participative policy-making evaluation does not seem to play as self-evident a role as it does in relation to NPM innovations. Still, many evaluations take place. In case of public-private co-production assessments are made to verify whether contractual the parties involved have met obligations and other arrangements. Also when projects run over their budgets or time schedules to a considerable extent, controlling bodies often undertake evaluation studies to assess how these outcomes came about and who is to be held responsible.

In cases of participatory policy processes participative evaluations (Cousins and Earl 1992) or other forms of multi-actor evaluations (Abma 1996; Taket and White 1997; Ryan and DeStefano 2001) may be done. Such evaluations, in which assessments from different actor perspectives find a place, are primarily meant to support further debate, development of new options and learning processes (Cf. Teisman and Van der Meer 2002).

It is my impression [to be tested] that in connection to these second and third types of public management innovations, there is little systematic evaluation of these innovations as such: in what way and to what extent are policy processes and policy outcomes improved by co-production and participation?

## **Dynamics of evaluation**

Before turning to the question how evaluation in and of public management can be improved, I venture a theoretical reflection about why evaluation practices are shaped the way they are, when, why and how evaluation results are used and how their impacts are generated.

I propose to view evaluation as a sensemaking process. It is about assessing what is or has been going on, how this has come about, how the findings have to be valued and perhaps about how we can think about possible improvements. Such sensemaking takes place in the context of already existing frames of meaning. Existing notions, perspectives and 'theories' are used to orchestrate the search for answers to new questions and interpret resulting new data. Thus, it is to be expected that the setup of evaluations in the framework of a specific management strategy or reform will reflect the dominant rationale behind this strategy or reform.

For example, evaluations related to NPM-type innovations will tend to start from the idea a result-oriented structure will evoke maximal motivation and productivity (both quantitative and qualitative). Therefore evaluation should assess results and measure these in terms of success or accordance to preset goals. The agent should then be rewarded accordingly. Within the NPM rationale this argument with respect to evaluation seems to be self-evident. From this perspective there is little reason to put the rationale itself to question, which may explain why NPM reforms are seldom explicitly evaluated. And when they are, evaluations tend to focus on the extent to which organizations involved bring NPM procedures and provisions into practice and far less on what the eventual impacts of the reforms are.

A second step in the theoretical argument is about the impact of evaluations (Cf. Van der Meer 1999). Here again the notion of sensemaking plays a key role. The influence of evaluations and evaluation results on actual behavior (talking, rewarding, changing, acting) is determined by how they are interpreted and valued by actors involved. It is important to note that in general there are far more actors involved than those who commissioned the evaluation. Although the latter may sometimes have the power to make decisions with respect to tasks, personnel, budgets, rewards, or organizational structures and procedures, they are by no means the only actors whose sensemaking matters. I have already pointed to the possibility that evaluation evoke perverse effects or defensive reactions if that seems to be useful or self-evident from the perspective of evaluated actors<sup>2</sup>. Moreover, evaluated actors may try to 'improve' their actions in anticipation or reaction on evaluation. The way in which they proceed in this connection will again depend on how they interpret both their own behavior and the evaluation outcomes.

Professional employees in a unit involved in some sort of contract may – from their professional perspective - make a rather sharp distinction between the quantified output criteria mentioned in the contract and 'real' quality, which is hard to measure. If so they may consider it fully legitimate to score on the output criteria as quickly and artificially as possible in order to save time for the real thing. Or they may view their professional position as strong enough to simply negate the output criteria and focus on the own norms.

A following theoretical idea is that third actors play a role. Evaluation is not a sensemaking process that takes place in isolation between principal and agent or between evaluator and evaluated, but in a context full other actors, such as competitors, clients, citizens, interest groups, parliament, press, etc. Such actors may formally or informally influence the focus or criteria use in evaluations. They may also react to evaluations (or their divers follow-ups). If these third actors (are considered to) have power or are thought of as 'important' by the focal actors, their reactions may contribute to an indirect influence evaluations may have on these focal actors. Moreover third actors may adapt their own behavior on the basis of their interpretation of evaluation outcomes.

Suppose, some targets are set for the personal safety policy in a certain neighborhood and that an evaluation report shows that safety has increased, while at the same time many inhabitants of the neighborhood feel less secure than they did a year ago. They may protest and seek media coverage for their problems. Or they may realize that they can influence (i.e. make more realistic from their point of view) performance indicators, e.g. by reporting incidents more frequently to the police. The first reaction may for example trigger a decision to include subjective experience of inhabitants in next year's performance review (as is actually don in the Rotterdam safety monitor). The second reaction may lead to a lower score in the following year, which may trigger measures that are not actually warranted by worse performance.

---

<sup>2</sup> In fact, the labels 'perverse' and 'defensive' are interpretations and valuations based on sensemaking processes themselves. There are few actors considering themselves perverse or defensive. This precisely exemplifies the diversity of sensemaking in these processes.

Many if not most policy processes are multi-actor in nature in the sense that the eventual policy is formulated or negotiated in interaction and co-operation between different administrative bodies and layers, and often also private and societal organizations. The same holds for many 'implementation' processes. In these cases, what has been said about third actors above applies in a specific way. Like citizens, clients and media co-producers can influence evaluation and will react to their outcomes (or not) and thus co-determine the setup, conclusions and impact of evaluations. More specifically, in cases of co-production – and perhaps also in some cases of participatory policymaking – it is not a single given actor who defines and commissions evaluations. Different actors may do so at the same time, thus producing concurring evaluations with potentially contradictory results, which will not necessarily contribute to substantive improvements or better cooperation. An interesting aspect in this connection is the question who initiates an evaluation, along which lines and especially: who is evaluated. Even apart from the substantive conclusion of an evaluation these elements may strongly influence sensemaking processes of the different actors and hence their reactions.

Dutch spatial developments actually are a co-production of different governmental bodies, private enterprises, environmental and other interest groups and citizens. If the central government or the Spatial Planning Bureau commissions an evaluation to assess to what extent the general goals formulated in the Spatial Policy Bill are realized, municipalities may experience this as an effort to increase central control, especially if they are not given the opportunity to insert their goals, limitations and assessments in the evaluation process (see Teisman et al. 2002, for a detailed analysis and some options for dealing with these complications).

This theoretical argument gives rise to the presupposition that the design, the outcome and the impact of evaluation are products of complex interaction processes in which sensemaking from a diversity of perspectives (I call these 'repertoires' (Van der Meer 1999; Van der Meer et al. 2000)) takes place. Then, insight in the relevant repertoires and knowledge of the interaction patterns between relevant actors helps to understand the dynamics around evaluations and their impacts. Based on such understanding ideas for more effective evaluations and evaluation arrangements may be developed.

### **Evaluation of public management reforms**

Above, I hypothesized that the relative lack of comprehensive evaluations of public management reforms is due to the self-evidence their rationales have for their promoters. Of course, from an 'external' perspective, such as developed in this paper, evaluation of public management reforms as such is very desirable. In this section I develop some ideas about what such evaluation might look like. The following section is devoted to the question how evaluations of public management innovations – and also evaluations within the framework of (new) public management practices - can be done in a way that enhances its utilization of evaluation results and their impact. For both themes I make use of the theoretical argument developed above.

If evaluation can be considered a sensemaking process in interaction in which actors apply their existing repertoires, we can as well presuppose that public management innovations 'work' via sensemaking and related behavioral adaptations by actors involved. This would imply that the meaning of such innovations may differ between actors and that their impact and 'success' is context dependent. To phrase it differently: the effects of an innovation do not only depend on characteristics of that innovation, but also – and primarily! – on the meaning it acquires from actors' perspectives and on the interplay of interpretation related actor behaviors. Thus, a typical evaluation question should not be whether new arrangement X is

‘good’, or ‘better’ than arrangement Y, but how arrangement X works in a specific context<sup>3</sup> and why. Evaluations should indeed search for explanations of the functioning and effects of public management innovations since this may provide new input in sensemaking and learning processes, contributing to ideas for improvement. Moreover, the analytical argument can potentially be generalized to other settings, which is generally not the case for substantive findings.

The theoretical approach advanced in this paper also implies that knowledge of the perspectives (repertoires) of actors involved (in whatever role) as well as of their interactions is necessary for explanations to be found. Therefore, interviews with actors in different roles and positions – or other forms of participation in the evaluation process – are required to reconstruct their repertoires and be able to explain their (re)actions and thereby the functioning and eventual effects of the management structure or strategy under review. It should be noted that this thesis is not based on a normative choice for democracy or empowerment (however honorable such a choice would be) but on an analytical argument on the dynamics of evaluation and public management innovations.

It should also be noted that involvement of stakeholders in the evaluation process may influence and enhance the impact of the evaluation since participants are more aware of the fact that there is an evaluation, may be more inclined to acquire knowledge on its content and conclusions, may recognize more elements, and may feel more committed to act upon the evaluation outcomes. In fact, resulting changes in actor sensemaking and behavior may also influence the functioning and effects of the public management innovation(s) under scrutiny. The direction of these changes depends on how evaluation results and other experiences in the evaluation process are linked by actors to their existing ideas and practices (Van der Meer 1999). That’s why the idea that participation produces support seems to be too simple, both with respect to the evaluation process and with respect to the public management innovation evaluated. It is quite conceivable that a negative attitude towards the public management innovation is reinforced or induced in the process for some actors. And even if actor attitudes change in a positive direction the question remains exactly what it is that he supports: his (changed) image of what the innovation is about is decisive.

Thus, although I have argued that some form of interaction with stakeholders in the evaluation is required for sensible evaluation of public management innovations, it should be realized that in doing so the object of investigation might change. Or, to put it differently, the evaluation process becomes part of the functioning of the innovation.

## **Towards effective evaluation**

The ideas on the dynamics of evaluation outlined above, also give cues for thinking about improving the efficacy of evaluations, both of and in public management innovations. The first question here is: how can we define and assess the effectiveness of evaluations. A number of levels of impact come to mind (Cf. Hupe and Meer 2002, 14):

1. actors evaluated and other stakeholders take note of the evaluation
  2. actors use evaluation results in considering changes in policy, management, working processes, etc.
  3. actors support the conclusions of an evaluation
  4. actual changes in policy (implementation) or management as suggested by the evaluation
- However, on reflection this scheme is problematic.

First, although condition 1. seems to be valid on the level of the network as a whole – it is a necessary condition for any impact to materialize – it is not so at the level of individual actors. Via chains of interactions in the network the behavior of some actors may change, partly as a consequence of the evaluation, without these actors having any knowledge of the evaluation (indirect impact).

---

<sup>3</sup> What the specific context is, is not trivial either. Actors may have different perceptions of or convictions about what the relevant context is.

Second, level 4 may come about as a consequence of other processes than the evaluation. In fact it is generally very difficult to attribute actual changes to a single cause or even to estimate the contribution of a give cause.

Finally, levels 3 and 4 are not necessarily higher or better than level 2. Criteria used, conclusion drawn and recommendations given in an evaluation reflect, as I suggested above, a specific perspective or repertoire. Thus, evaluation results or other experiences in the evaluation process may acquire different meanings for other actors, making sense utilizing their own repertoire. Although the evaluators may judge this negatively, from any other stance it may be considered a potential contribution to new ideas, new innovations and hence learning

Therefore, my thesis is that the only *general* criterion for effectiveness of evaluation is the extent to which it contributes to learning processes in the network of relevant actors/stakeholders. The extent to which this is the case can be related to the number or the diversity of actors learning (or reporting to learn) something from the evaluation, or to the measure to which new concepts or 'facts' from the evaluation are used in consecutive discourse. 'Learning' might also be measured by the extent to which actors attribute actual changes (especially in their own behavior) to the evaluation or to concepts and 'facts' produced by the evaluation. Finally, learning can be measured by the extent to which evaluations trigger interaction and debate in the network.

One could point to the fact that many evaluations, especially in the context of NPM, do not primarily aim at learning, but rather at control and accountability (see Sanderson 2001, on this distinction; Teisman et al. 2002). However true that may be, my argument would be that the efficacy of accountability oriented evaluation depends on the extent to which and the way in which it is include in sensemaking processes of and among actors involved. If no actor reads something new in the evaluation, if its results are no object of debate and if no changes in behavior can be related to the evaluation, the evaluation has not been effective, since no learning took place.

Thus, the key question becomes how learning from evaluation can be enhanced.

The theoretical argument in this paper suggests that actors should both recognize the evaluation (outcomes) as sensible from their existing repertoire and as new. If the first condition is not met, the evaluation is simply noise to them, if the second is not met, there is nothing to be learned (except the fact that there is nothing to be learned, which consolidates existing repertoires and may decrease opportunities for future learning). The 'connectivity' of evaluations – the extent to which they are connectable to existing repertoires of different actors, thus inserting new element in these – depends on both their contents and the interaction processes in which they are embedded. Actors commissioning or performing evaluations can enhance connectivity by gaining knowledge of actor repertoires, related preoccupations, dominant questions etc., to which the evaluation design can be adapted. Alternatively they can involve second and third actors in the process of designing the evaluation.

It should be realized in this connection that the preconditions for connectivity and learning are generally different for different actors, since their repertoires and positions in interaction patterns differ. By consequence, it will not be possible to design evaluations that maximize learning for all actors at the same time. Actors have different interest and questions. Especially in the case of co-productive or participative policy processes, actors should be allowed, enable and stimulated to perform or commission their own evaluations, thus feeding their own learning processes. This may also reduce perverse behavior or defensive reaction (however these may be defined).

But it should also be realized that policy and management results eventually are produced in the interplay between all actors. If evaluation is to stimulate and improve their cooperation, it should also help their mutual debate and collective learning processes. Efforts should therefore be undertaken to enable linking of and confrontation between different evaluations, done by different actors at different levels, and to supplement them with joint (multi-actor) evaluations (Teisman et al. 2002; Meer and Edelenbos to be published). In the latter case collective learning and mutual cooperation can already take place in the design phase of the evaluation.



## Conclusions

In this paper I dealt with some issues concerning evaluation of and in public management innovations. I proposed an interactionist of constructivist theoretical framework to analyze the dynamics of evaluation. By conceiving of evaluation as a sensemaking process in interaction between actors in a network, both the design and the impact of evaluations can be explained (as well as the lack of evaluation or the lack of impact). Moreover the argument gave rise to some ideas about shaping evaluations of public management innovations as such. These can be summarize as follows:

- the success and impacts of public management innovations are not only determined by characteristics of the innovation; evaluations should therefore be contextualized
- evaluations of public management innovations should include (context dependent) explanations for their functioning and effects, both for context related learning and for generalization to other contexts
- these explanations require knowledge of actor repertoires, which implies some sort of participation of key actors and stakeholders in the evaluation process
- evaluation of public management innovations influence the functioning and impact of these innovation, i.e. evaluation becomes part of the innovation

Finally I derived some conclusions with respect to enhancing efficacy of evaluations (both of and in public management innovations):

- learning is the only general criterion for evaluation effectiveness
- connectivity of evaluations in relation to actor repertoires determines their learning potential
- since different actors have different repertoires and learning needs, a multiplicity of evaluations should be favored performed by different actors, at different levels and with different core questions. However these evaluations should have enough connectivity to contribute to mutual debate and cooperation. Also multi-actor evaluations should be an element of the set of evaluations in a network.

Further research is needed to test the theoretical notion advance in this paper, especially their application to public management innovations and related evaluations. Also the recommendations should be tested and evaluated in their own right. For now, however, I think these ideas and arguments warrant some discussion and critique.

## References [further references to be added]

Abma, T. A. (1996). *Responsief evalueren: discourses, controversen en allianties in het postmoderne*. Rotterdam.

Cousins, J. B. and L. M. Earl (1992). "The case for participatory evaluation." *Educational Evaluation and Policy Analysis* **14**(4): 397-418.

Edelenbos, J. (2000). *Proces in vorm: procesbegeleiding van interactieve beleidsvorming over lokale ruimtelijke projecten*. Utrecht, Lemma.

- Forbes, M. and L. E. L. Jr. (2005). "How does public management affect government performance? Findings from international research." *Journal of Public Administration Research and Theory* **15**: 559-584.
- Hupe, P. and F.-B. v. d. Meer (2002). Doorwerking van emancipatie-effectrapportages in beleidsprocessen. Den Haag, Ministerie van Sociale Zaken en Werkgelegenheid.
- Kickert, W. J. M., Ed. (1997). *Public management and administrative reform in Western Europe*. Cheltenham, Northampton, Edward Elgar.
- Meer, F.-B. v. d. and J. Edelenbos (to be published). "Evaluation in multi-actor policy processes: accountability, learning and cooperation." *Evaluation*.
- Meyer, M. W. and V. Gupta (1994). "The performance paradox." *Research in Organizational Behavior* **16**: 309-369.
- Pollitt, C. (1995). "Justification by works or by faith? evaluating the new public management." *Evaluation* **1**(2): 133-154.
- Pollitt, C. and G. Bouckaert (2000). *Public management reform: a comparative analysis*. Oxford, Oxford University Press.
- Pollitt, C. and G. Bouckaert (2003). Evaluation in public sector reforms: an international perspective. *Evaluation in public sector reform*. H. Wollmann. Cheltenham, Northampton (MA), Edward Elgar: 12-35.
- Ryan, K. and L. DeStefano (2001). "Dialogue as a democratizing evaluation method." *Evaluation* **7**(2): 188-203.
- Sanderson, I. (2001). "Performance management, evaluation and learning in 'modern' local government." *Public Administration* **79**(2): 297-313.
- Taket, A. and L. White (1997). "Working with heterogeneity: a pluralist strategy for evaluation." *Systems research and behavioral science* **14**(2): 101-111.
- Teisman, G. R. and F.-B. Van der Meer (2002). Evalueren om te leren: naar een evaluatiearrangement voor de Vijfde Nota RO. Rotterdam, Erasmus Universiteit.
- Teisman, G. R. and T. J. M. Verheij (1996). Draagvlakvorming bij technisch-complexe projecten. *Grote projecten*. J. A. d. Bruijn, P. d. Jong, A. F. A. Korsten and W.P.C. van Zanten. Alphen aan den Rijn, Samson H.D. Tjeenk Willink: 174-192.
- Tops, P. W. and R. Weterings (1998). Gemeentelijk beleid en co-productie. *Lokaal bestuur in Nederland, inleiding in de gemeentekunde*. A. F. A. Korsten and P. W. Tops. Alphen aan den Rijn, Samsom: 518 - 528.
- Van der Meer, F. B. (1999). "Evaluation and the social construction of impacts." *Evaluation* **5**(4): 387-406.
- Van der Meer, F. B., G. J. D. De Vries and G. A. N. Vissers (2000). "Evaluatie en leerprocessen bij de overheid: de rol van institutionele condities." *Beleidswetenschappen* **14**(3): 253-277.