

Evaluation and NPM – as seen from the perspective of evaluation

Peter Dahler-Larsen | *In an era of New Public Management (NPM), it is important to discuss the role of evaluation. In order to do so, assumptions need to be unpacked. Neither NPM nor evaluation are fixed terms. Borrowing a substitution hypothesis and a complementarity hypothesis from Balthasar and Rieder (2009), this article discusses the role of evaluation in these two scenarios and suggests that evaluation may learn from and respond to a situation where it is put in a defensive role. However, a more offensive role is also possible. Critical discussions of NPM take place in many countries, so NPM is not historically inevitable. Evaluation is much more than a management instrument, so a symmetric comparison between NPM and evaluation may be unfair. Beyond NPM, evaluation may have important roles to play in a modern, democratic society.*

Table of contents

- 1 Introduction
- 2 The meaning of the term 'evaluation'
- 3 The meaning of the term 'New Public Management'
- 4 The relationship between evaluation and NPM: The substitution hypothesis
- 5 The relationship between evaluation and NPM: The complementarity hypothesis
- 6 NPM is not historically inevitable
- 7 Looking forward: Evaluation beyond NPM

1 Introduction

Does New Public Management seriously threaten the role and vitality of evaluation? To the best of my understanding, this was the question that motivated the organizers of the SEVAL conference in Fribourg in 2013 to invite me to speak on the basis of the headline of this article, an invitation that I gratefully accepted. I shall speak in the capacity of a teacher of evaluation, a friendly, but critical researcher in evaluation, and as a former president of the European Evaluation Society. I have been asked to talk about the issue “as seen from the perspective of evaluation”. It is thus my role to interpret our situation in such a way that the positive potential in evaluation can be identified, as long as I remain within the limits of sound conceptual work and reasonable critical thinking.

The question concerning the relationship between evaluation and NPM deserves to be raised in any country where an evaluation society has been established in order to enhance the role and quality of evaluation in public affairs, and where, at the same time, the standards, procedures and mentalities of New Public Management become not only an important model, but perhaps *the* dominant model for public organization, decision-making and change.

Since its very inception, evaluation has been in interaction with various tendencies, ideas and models of organization in society. It is not new for evaluation to change, adapt and enter into dialogue with the surrounding society. In fact it is characteristic of evaluation to do so. New Public Management, however, is a strong idea in many national and international contexts, and it is so functionally close – and thus in competition with – many aspects of evaluation that many evaluators may fear for their future and the future of their field.

However, fear should not cloud analytical thinking, nor should it prevent us from seeing the future as open and full of options. As always, when social scientists are asked to make predictions, we should remind ourselves that the future is a construction that depends on collective action. Predictions are better thought of as scenarios that motivate us to act rather than as descriptions that are only valid if they are precise.

In this light, I shall seek to demonstrate that our answer to the question about the relationship between evaluation and New Public Management depends on assumptions that are variable. First, I shall discuss the meaning of the terms ‘evaluation’ and ‘New Public Management’. Second, I shall argue that we should not focus too much on the ‘amount’ of evaluation and NPM respectively, but instead on their relationship to one other. Here I shall build on a distinction between a substitution hypothesis and a complementarity hypothesis borrowed from Balthasar and Rieder (2009). Third, I will argue that NPM has already shown inherent weaknesses in many countries and that its dominance is less historically inevitable than is often assumed. Fourth, I will argue that evaluation has a broad spectrum of functions and roles. It is more than a management instrument. In this light, a symmetrical comparison between New Public Management and evaluation is not fair.

I find it important to unpack these reservations step by step. Assumptions are not only building blocks in analysis that may be more or less sound. They are also social and political constructions. My point is that if particular assumptions about evaluation and NPM are kept in place by institutions and by people in a particular socio-political context, then these assumptions may in fact influence the real future of evaluation and NPM in that context. In order for evaluation to fulfil its potential, it may be necessary to break with certain assumptions that some people have about evaluation.

In other words, the present article does not aspire to deliver a good prediction of whether NPM will replace evaluation. Instead, it provides bits and pieces of a theory that helps us understand the forces and ideas that shape evaluation, NPM, and the relationship between the two.

2 The meaning of the term ‘evaluation’

Evaluation is not only an activity, it is also a field consisting of ideas, models, perspectives, theories and experiences. This field has so far demonstrated quite an impressive capacity to change, diversify and renew itself. New evaluation models have constantly been developed to compensate for the weaknesses of older ones.

While all evaluation somehow operates within a four-dimensional conceptual space consisting of values, methods, evaluands, and utilization (Shadish et al 1991), there is, within that space, a great variety of models and approaches that serve different needs. Although many evaluators have their favorite evaluation models, most theorists of evaluation believe that a situation analysis should be conducted before each evaluation in order to ensure the optimum match between the situation and the chosen model (or models). However, a situation analysis depends on an *interpretation* of the situation (Dahler-Larsen and Schwandt 2012). Evaluation is not only a passive instrument, it is in fact an active player that – based on a view of the evaluand – sets in motion a particular configuration of values, methods and intended use of the evaluation.

If we conceive of a democratic society as one which attempts to master its own destiny, then evaluation is born with a clear democratic mandate. Evaluation is based on the belief that new and systematic knowledge about initiatives of common interest in society can be used to improve such initiatives in the future. Evaluation is a part of the self-appropriation of a democratic society (Rosanvallon 2009; Stehr 2001). All of the key terms involved here, ‘systematic’, ‘knowledge’, ‘common interest’, ‘society’, and ‘improve’ are fairly open and contested, as they should be in a democracy. In that sense, evaluation is an unfinished social construction.

For that reason, evaluation can only be provisionally described in terms of the conceptual space in which it operates, but cannot be finally determined as one, and only one form of practice. In this light, it makes perfect sense that the field of evaluation has developed new approaches and models on an ongoing basis. If we are to define evaluation more specifically, we are reminded of Reinhard Kosellecks (1972) words about the “living tension between reality and concept”.

While evaluation, of course, takes on a specific shape and form in any given socio-historical situation, this particular shape and form should not be confused with evaluation as such.

3 The meaning of the term ‘New Public Management’

Most definitions of New Public Management (NPM) include a focus on indicators which describe the outputs and outcomes of public organizations. These indica-

tors are believed to be an important instrument in the hands of public managers in order to improve services and make them more efficient, but in a larger perspective, measurement of outputs and outcomes also paves the way for a broader restructuring of relations between various organizational units. These relations can be more clearly contract-based (rather than bureaucratic) depending on whether or not we are looking at relations between different branches or levels of public administration, or at relations between governmental agencies and private contractors.

While these ideas seem to constitute a more or less common ground in ideas related to New Public Management, there is great variation with regard to their implementation and their institutionally defined consequences, including the sometimes limited effect of performance measurement upon decision-making (Pollitt 1995). In fact, the literature suggests that NPM is not only multifaceted (Pollitt 1995), but actually consists of more than a handful of components, such as contracting-out, a client-centered view of quality, benchmarks and quality standards, and financial reforms (Ferlie and Steane 2002). Diefenbach (2009) identifies elements of NPM under the following categories: business environment and strategic objectives, organizational structures and processes, performance management and measurement systems, management and managers, and employees and corporate culture. Taking this composite view of NPM one step further, I will argue that the same configuration of components is not always mobilized in every instance. In fact it is meaningful to speak of at least three different orientations in NPM.

According to the first of these, NPM is primarily about how to *think* about public services. The idea is that service-providers are there for the citizens. Indicators can be used to express new and better visions of the difference that services make to citizens. That is why outputs and – better still – *outcomes* for citizens should be measured. If successful, indicators help dismantle a bureaucratic mentality and focus on the needs of citizens. Here the primary use of indicators is *motivational*.

According to another orientation, NPM has to do with what is *known* about the public sector. Public management should be based on exact knowledge about the effects of interventions, about citizen satisfaction with services etc. More efficient production of public services requires systematic knowledge. While the primary use of this knowledge may be managerial, it is also possible that service users, as well as politicians and indeed other stakeholders in public life may benefit from the publication of knowledge about the effectiveness and efficiency of public services. Here the primary use of indicators is *informational*.

The third orientation in NPM looks at how public services are paid for. Results should be rewarded. Indicators are linked to money flows. In this variation of NPM, the financial framework of a given public institution or service provider depends on its score on relevant indicators of outputs or outcomes. Here, contractual relations can cease to exist if they are not seen as efficient. This strong version adds real money to the restructuring forces of NPM. The primary use of indicators lies in *incentive structures*.

Each of these three orientations may embody both theoretical and practical weakness, such as how to measure outcomes, but that is not our point here. The point is that while all of these three orientations operate under the umbrella of NPM, they are fundamentally different in terms of their implications and consequences. It may therefore be tempting for advocates of NPM to rhetorically emphasize the benefits of several of the versions, while sticking to other versions or the general idea when it comes to difficulties, costs, and downsides.

In other words, although it may be true that NPM has moved forward in many countries, we should acknowledge that it may be implemented in many different ways, and the specific form it takes in a particular politico-administrative context may be neither stable, consistent nor totally in line with official rhetoric about NPM.

4 The relationship between evaluation and NPM: The substitution hypothesis

As explained above, both evaluation and NPM may be more flexible, more multifaceted, more dynamic and less unequivocal than is often assumed. With this observation in mind, we shall now look more closely at the relationship between the two.

One hypothesis, the substitution hypothesis, suggests that over time NPM will gradually take over the role of evaluation. At least three theoretical ideas lend support to this hypothesis. First, attention is limited in organizations (Simon 1969), and key people in public organizations are rarely occupied by more than a few key ideas at the same time. Fashionable ideas come and go in a way that does not include a careful assessment of all pros and cons of each idea (Røvik 2005).

Second, dominant ideas are often carried by institutions. The implication is that if powerful institutions (for example, the national audit office, the Ministry of Finance, and dominant consulting companies) adopt the mentality of NPM, they can use their position to promote the idea. International organizations, such as the OECD and the EU, have also provided institutional support for the circulation of ideas about the structuration of governance and public management. (I shall leave the analysis of the specific historical and political case of

Switzerland to others with the required expertise). The general point, however, is that diverse ideas require institutional protection, and dominant ideas often reside in dominant institutions.

Third, forms of organization depend on their social circumstances and larger social imaginaries (DiMaggio and Powell 1983). One way of analyzing these larger social imaginaries is by means of a distinction between reflexive modernization (Beck 1992) and the audit society (Power 1997). Reflexive modernization is an era in which the side effects of modernity become visible. Reorganization, globalization, technology and IT help create social contingency, leading to a mentality of ongoing change. Cultural diversity helps enhance multiple perspectives. In a situation where progress based on rational consensus is no longer possible, only an abstract myth of development can guide the social imaginary. The myth of development means taking side-effects and multiple perspectives into account on an ongoing basis. The field of evaluation has developed a number of evaluation approaches that respond to this social imaginary, such as responsive, participatory, deliberative, and transformative evaluation (Greene 1997; Stake 2004).

In recent years, we have seen the advent of the audit society (Power 1997). Its dominant myth is one of assurance. In the audit society, activities need to be checked *before* they go wrong. Comprehensive, mandatory surveillance and reporting mechanisms are put in place not to create improvement and development, but to manage risk in complex systems. This is the era of evaluation machines based on handbooks, guidelines, reports, inspection, and indicators. The mentality of the audit society fits into a broader social imagery that since 2001 has been occupied with the avoidance of risk and disaster.

If it is sociologically correct that the mentality of the audit society has gained ground, then this observation lends theoretical support to the substitution hypothesis, but we can add subtle nuances. The transformation in favor of surveillance and indicators is not only due to the strength of new ideas, but also to weaknesses in older ones. The very field of evaluation (when dominated by the spirit of reflexive modernization) has namely been criticized for not being reliable, for being too subjective, for producing many evaluations that did not lead to synthesis and managerial overview, and for being of too little use (Dahler-Larsen 2012). The recent interest in evaluation capacity development, in evaluation policy and evaluation culture testifies to the relevance of this critique and to the relevance of responding to it.

In other words, if the substitution hypothesis is correct, evaluation is forced into a defensive position, but it is not left without options. If evaluation can become more systematic, more organizationally relevant, and enjoys more widespread use, then it is at least able to defend itself against the criticism

levelled against it by the audit society. If NPM poses a threat, evaluation can learn from this threat. Evaluators should not predict, they should respond.

5 The relationship between evaluation and NPM: The complementarity hypothesis

According to another hypothesis, the complementarity hypothesis, there is a place and a role for both NPM and evaluation. We have several ideas that would support such complementarity. If a careful systematic analysis is made before information is collected, there would sometimes be a need for ongoing performance management, and sometimes a need for custom-tailored evaluation. The production of knowledge would take the form of an ongoing dialogue between questions and answers. For example, some would focus on results, others on processes. Some would require quantitative and qualitative information, respectively.

It would be possible to conceive of a systematic evaluation system, evaluation policy or evaluation portfolio that describes a division of labour between performance indicators and evaluation. There could be functional divisions and varying intensity over time in the different approaches.

Evaluation could complement NPM in a number of ways. Evaluation could identify implementation problems as the paraphernalia of indicator systems is not strong with respect to process analysis. In a similar vein, evaluation could help identify causal links in processes between interventions and their outcomes (Pawson and Tilley 1997). Evaluation could also help qualify success criteria and indicators through the use of good program theory and/or by testing the relevance of criteria for users and other stakeholders. These would be roles for evaluation that are likely to be accepted by most NPM advocates. We could also imagine two more roles for evaluation, still compatible with the complementarity hypothesis, but this time with a more dialogic and perhaps competitive role between evaluation and NPM rather than an agreed-upon division of labour.

I am here thinking of evaluation as an approach in the meta-evaluation of NPM reforms. It was quite a number of years ago that Pollitt (1995) argued that NPM reforms, despite the official claims of transparency and effectiveness, were rarely evaluated systematically. Herein lies an important mission for evaluation in an era otherwise dominated by NPM.

Furthermore, the field of evaluation provides a theoretical space for reflection about the choice of values, the choice of methods, and the use of knowledge in the public sector. Within this theoretical space, it is actually possible to conceive of indicator systems as a form, although a very narrow and specific form, of evaluation. Students trained to think evaluatively within such conceptual space may, as professionals, begin to question indicator systems. Why this indicator rather

than another? How can this indicator be justified based on a theory of values? What are our assumptions about the use of indicators?

In this view, evaluation is not only a practice, but a field and a conceptual set of ideas that deserves to have a place, sometimes as a helper of NPM, and sometimes as a critical friend. Even if the complementarity thesis is correct, evaluation may play more than one role in relation to NPM.

6 NPM is not historically inevitable

If our question about the relationship between NPM and evaluation and our remarks about the “era of New Public Management” has been based on an idea about automatic progress of NPM, any assumption about such historical inevitability should be questioned. Like evaluation in general, NPM is contingent upon a broad set of societal forces (Dahler-Larsen 2006; 2012).

In fact, there is enough experience with NPM in some countries to reveal serious points of contention and discussion.

A major issue has to do with confusion regarding results, outcomes and effects. While NPM claims to be results oriented, the measurement and management of results has proven to be difficult. The results of public services are sometimes difficult to measure unless their meaning is transformed or reduced. For example, although schools do have results, the meaning of education in society remains broad and is sometimes philosophically, rather than statistically defined. Furthermore, there is conceptual confusion between results, outputs, outcomes, impacts and effects. Although, of course, there are attempts to confine each of these concepts to a clear definite meaning, more often than not they are confused in practical-political and managerial discourse. In evaluation theory, as well as in social science thinking, clear attempts are made to distinguish between situations where a causal link is assumed or demonstrated, and situations where it is not. For example, Vedung (1997) reserves the term ‘effect’ strictly for the former situation, while a more free-floating terminology may be permitted for the latter. The crucial point is that if a causal link is assumed, i.e. a particular situation occurred because of a previous public intervention, then it needs to be taken seriously whether it is methodologically possible to demonstrate such a link. It is not enough to measure a variable and call it an outcome.

It is easy to say that something was a result of something else, but without serious attention to the methodological validity of claims that support the statement, it is easy to overestimate or underestimate the merits of public interventions. Various schools of thought and various evaluation manoeuvres such as experimental designs or statistical controls help identify and narrow down the size of any causal effect.

Often, however, indicator systems built in the name of NPM do not always meet tough methodological requirements. A theory of causal links may be missing, good indicators may be missing, and most seriously, there is no genuine control for other factors that influence a particular problem situation in society. Even if control factors are added into the equation, such as socioeconomic background factors that influence test results in schools, there is often still a wide range of scores that are statistically possible for any given school because even a wide set of predictors only predict school results with a certain degree of statistical uncertainty. This is well known among statisticians. However, if causal attribution is made only in imperfect ways, the social, political and managerial implications are controversial. To what extent should a public institution accept responsibility and blame for particular scores that are, at best, statistically uncertain? To what extent can financial and even legal consequences be connected to causal analyses that are, again, at best uncertain? If a performance management regime or inspection regime that includes causal claims leads to sanctions, to what extent is uncertainty in that causal analysis legally, morally and administratively acceptable? Democratic societies usually issue legal sanctions only if based upon knowledge beyond reasonable doubt. Most social scientists admit, however, that reasonable doubt is a sound ingredient in most causal analyses in practice.

In the event of a lack of trustworthy causal evidence, performance management indicators often go one step back in the production chain and focus only on output measures. How much was produced of this and that, how many operations were carried out, how many meals were sent out, how many teeth were brushed? The figures resulting from this exercise sometimes remind one more of a classic five-year plan and production measures than a genuine focus on outcomes relevant for the well-being of citizens.

Sometimes, an intense focus on holding particular public organizations accountable for a small set of indicators leads to a defensive view of quality. Furthermore, this type of knowledge production does not always lead to increased knowledge about which types of interventions work. There is a trade-off between evaluative knowledge that connects results to productive units and evaluation knowledge connected to types of activities. The main emphasis in NPM may lie in the former at the expense of the latter.

However, as NPM as an ideology helps institutionalize routine-based indicator systems, people may behave as if a good score on these indicators is a goal in itself. Constitutive effects may occur. Indicators may help to create new realities. Constitutive effects is a concept that reaches well beyond the distinction between intended and unintended effects. While some effects of indicators may truly be unintended, it is not enough just to use some set of well-defined original inten-

tions as a contrast. In complex administrative systems, there are many actors, and they sometimes invent new intentions along the way. They use indicator systems for various purposes along the way, and the purposes are many (Behn 2003). Various side effects occur, and compensations for these side effects are sometimes built into the indicator system, or are transported to other levels and agencies in the administrative system, for example decentral ones. However, if organizations at this level are given the task of measuring the unmeasurable, or are held responsible for the risk of low scores which they cannot control, then the dynamic life of an indicator system can be understood as a series of ongoing attempts to localize and re-localize the risks that are produced by the system itself (Rothstein, Huber and Gaskell 2006).

Faced with complexity – including the complexities of reflexive modernization which never disappeared but perhaps went “out of fashion” – NPM seeks to control the world through simplistic indicators that are, again simplistically, linked to particular agencies and organizations in order to hold them accountable. However, says the critique, this approach is inadequate as a response to the complex problems in today’s society. For example, to handle these problems, public organizations need much more horizontal coordination than is suggested by the accountability mentality inherent in NPM. Outlines of a post-NPM era can already be seen on the horizon.

7 Looking forward: Evaluation beyond NPM

Regardless of the specific status and position of NPM, evaluation has broader roles and potential functions in society that extend beyond being a management instrument that can be compared directly to NPM (Dahler-Larsen 2005, 616).

Evaluation can enhance enlightenment, debate and learning both inside and outside of managerial circles.

More specifically, evaluation has an important role to play with respect to the analysis of complex problems and complex interventions. We can already see interesting developments in the field of evaluation that seek to take into account both technical and social complexities (Funnell and Rogers 2011; Patton 2011).

The field of evaluation has traditionally paid attention to a broad set of values, and has been concerned with the issue of how values are justified, not only in relation to the definition of evaluation criteria, but also with respect to the evaluation process and the use of evaluation. As such, evaluation is a field and type of intellectual practice that has promising implications for other practices which are not called evaluation, but which are conceptually compatible, such as accreditation, auditing, benchmarking, and performance management. Even for the evaluation of NPM initiatives and reforms, evaluation is necessary per se. Last

but not least, evaluation has an important role to play in the public democratic debate.

This optimism in terms of the potential role of evaluation is not inconsistent with the observation that in some regards, evaluation is on the defensive vis-à-vis NPM. This situation, however, can be seen as an opportunity to learn, develop, and maybe even re-vitalize evaluation. Evaluation has done so before; in fact it has never rested for long, and there is no good reason why it should do so in the years to come.

Professor Peter Dahler-Larsen, PhD, Department of Political Science, University of Copenhagen, email: pdl@ifs.ku.dk

References

- Balthasar, Andreas / Rieder, Stefan, 2009, Wo ist evidenzbasierte Politik möglich? Die Verbreitung von Evaluationen auf kantonaler Ebene, in: Vatter, Adrian / Varone, Frédéric / Sager, Fritz (Hrsg.), *Demokratie als Leidenschaft*, Bern, Haupt, S. 403-432.
- Beck, Ulrich, 1992, *Risk Society: Towards a New Modernity*, London, Sage.
- Behn, Robert D., 2003, Why Measure Performance? Different Purposes Require Different Measures, *Public Administration Review* 63(5), S. 586-606.
- Dahler-Larsen, Peter, 2005, Evaluation and Public Management, in: Ferlie, E. / Lynn, L. / Pollitt, Christopher (Hrsg.), *The Oxford Handbook of Public Management*, Oxford, Oxford University Press, S. 615-642.
- Dahler-Larsen, Peter, 2006, Evaluation after Disenchantment: Five Issues Shaping the Role of Evaluation in Society, in: Shaw, I. / Green, J. / Mark, M. (Hrsg.), *The Sage Handbook of Evaluation*, London, Sage, S. 141-160.
- Dahler-Larsen, Peter, 2012, *The Evaluation Society*, Stanford, Stanford University Press.
- Dahler-Larsen, Peter / Schwandt, Thomas, A., 2012, Political Culture as Context for Evaluation, *New Directions for Evaluation*, Vol. 135, S. 75-88.
- Diefenbach, Thomas, 2009, New Public Management in Public Sector Organizations: The Dark Sides of Managerialistic "Enlightenment" *Public Administration* 87 (4), S. 892-909.
- DiMaggio, Paul J. / Powell, Walter W., 1991, The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields, in: Powell, Walter W. / DiMaggio, Paul J. (Hrsg.), *The New Institutionalism in Organizational Analysis*, Chicago, University of Chicago Press, S. 41-62.
- Ferlie, Ewan / Steane, Peter, 2002, Changing Developments in NPM, *International Journal of Public Administration* 25 (12), S. 1459-1469.
- Funnell, Sue C. / Rogers, Patricia J., 2011, *Purposeful Program Theory*, San Francisco, Wiley.
- Greene, Jennifer, 1997, Participatory Evaluation, in: Mabry, Linda (Hrsg.), *Evaluation and the Postmodern Dilemma*, London, JAI Press, S. 171-189.
- Koselleck, Reinhart, 1972, Begriffsgeschichte und Sozialgeschichte, in: Koselleck, Reinhart, 1979, *Vergangene Zukunft: Zur Semantik geschichtlicher Zeiten*, Frankfurt am Main, Suhrkamp.
- Patton, Michael Quinn, 2011, *Developmental Evaluation*, New York, Guilford Press.
- Pawson, Ray / Tilley, Nick, 1997, *Realistic Evaluation*, Thousand Oaks, Sage.
- Pollitt, Christopher, 1995, Justification by Works or by Faith? Evaluating the New Public Management, *Evaluation* 1 (2), S. 133-154.
- Pollitt, Christopher, 2006, Performance Information for Democracy: The Missing Link?, *Evaluation* 12 (1), S. 38-55.
- Power, Michael, 1997, *The Audit Society*, Oxford, Oxford University Press.
- Rosanvallon, Pierre, 2009, *Demokratin som problem*. [Democracy as a problem: "Inaugural lecture at Collège de France" and "Towards a Conceptual History of Politics"], Hægersten, Tankekraft.
- Rothstein, Henry / Huber, Michael / Gaskell, George, 2006, A Theory of Risk Colonization: The Spiralling Regulatory Logics of Societal and Institutional Risk, *Economy and Society* 35 (1), S. 91-112.
- Røvik, Kjell A., 2005, *Moderne Organisasjoner: Trender i Organisasjonstenkningen ved Tusenårskiftet*. Bergen, Fagbokforlaget.
- Simon, Herbert A., 1969, *The Sciences of the Artificial*, Massachusetts, MIT.
- Shadish, William R. / Cook, Thomas D. / Leviton, Laura C., 1991, *Foundations of Program Evaluation*, Newbury Park, Sage.

- Stake, Robert E., 2004, Standards-Based and Responsive Evaluation, Thousand Oaks, Sage.
- Stehr, Nico, 2001, The Fragility of Modern Societies: Knowledge and Risk in the Information Age, London, Sage.
- Vedung, Evert, 1997, Public Policy and Program Evaluation, New Brunswick, Transaction Publishers.

Zusammenfassung

Im Zeitalter des New Public Management (NPM) ist es wichtig, die Rolle der Evaluation zu diskutieren. Dazu müssen einige Prämissen erörtert werden. Weder «NPM» noch «Evaluation» sind fest definierte Begriffe. Unter Rückgriff auf die Substitutions- und die Komplementaritätshypothese nach Balthasar und Rieder (2009) analysiert dieser Beitrag die Rolle der Evaluation in diesen beiden Szenarien und regt an, dass die Evaluation in Situationen, wo sie in der Defensive ist, reagieren und aus diesen lernen kann. Es ist aber auch eine offensivere Rolle denkbar. In zahlreichen Ländern wird das NPM kritisch diskutiert, woran sich zeigt, dass dieses nicht historisch unvermeidlich ist. Da Evaluation viel mehr ist als ein Managementinstrument, wäre ein symmetrischer Vergleich zwischen NPM und Evaluation ungerecht. Die Evaluation kann über das NPM hinaus in einer modernen, demokratischen Gesellschaft wichtige Rollen spielen.