



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Eight attention points when evaluating large-scale public sector reforms

Hansen, Morten Balle; Breidahl, Karen Nielsen; Furubo, Jan-Eric; Hjørdis Halvorsen, Anne

Published in:
Handbook of Social Policy Evaluation

Publication date:
2017

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Hansen, M. B., Breidahl, K. N., Furubo, J-E., & Hjørdis Halvorsen, A. (2017). Eight attention points when evaluating large-scale public sector reforms. In *Handbook of Social Policy Evaluation* Edward Elgar Publishing.

General rights

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

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

Take down policy

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

Eight attention points when evaluating large-scale public sector reforms

Morten Balle Hansen, Department of Political Science, Aalborg University,
mbh@dps.aau.dk

Karen Nielsen Breidahl, Department of Political Science, Aalborg University,
knb@dps.aau.dk

Jan-Eric Furubo, Swedish National Audit Office, jan-eric.furubo@riksrevisionen.se

Anne Halvorsen, Head of Department, Department of Sociology and Social Work,
Agder University, anne.halvorsen@uia.no

Abstract

This chapter analyses the challenges related to evaluations of large-scale public sector reforms. It is based on a meta-evaluation of the evaluation of the reform of the Norwegian Labour Market and Welfare Administration (the NAV-reform) in Norway, which entailed both a significant reorganization of the central, regional and local government and a social policy reform. Meta-evaluations assess the usefulness of one or more evaluations and should not be confused with meta-analyses. The purpose of this meta-evaluation was to identify general principles for organizing the evaluations of large-scale public sector reforms.

Based on the analysis, eight crucial points of attention when evaluating large-scale public sector reforms are elaborated. We discuss their reasons and argue that other countries will face the same challenges and thus can learn from the experiences of Norway.

Preprint version – to be included in Handbook of Social Policy Evaluation:

Hansen, M. B., Breidahl, K. N., Furubo, J.-E., & Halvorsen, A. 2017. Eight attention points when evaluating large-scale public sector reforms. In B. Greve (Ed.), *Handbook of Social Policy Evaluation*. London: Edward Elgar Publishing.

Introduction

In this chapter, we explore what aspects require special attention when evaluating large-scale public sector reforms.

Our response to the question builds empirically on a meta-evaluation of the largest public-sector reform of the 2000s in Norway – the reform of the Norwegian Labour Market and Welfare Administration (called the NAV-reform in what follows), which was passed by the Norwegian Parliament in 2005. The NAV-reform entailed both a significant reorganization of the central, regional and local government and a social policy reform. When the reform was adopted, the Norwegian Parliament decided that it should be thoroughly evaluated. On this background, a large-scale research-based evaluation was organized, the major findings of which were presented at a conference in Oslo in June 2014 and in a final book from the evaluation (Andreassen & Aars, 2015). The evaluation has been subjected to a meta-evaluation (Stufflebeam, 2001) and it is the results of this that, in a condensed and worked up form, are presented in this chapter¹ (Breidahl, Furubo, Halvorsen, & Hansen, 2014). The purpose of this meta-evaluation has been to extract general experiences from the evaluation, which are relevant for the organization of future research-based evaluations of large-scale public sector reforms.

Theoretically, our response is based on the thesis that large-scale public sector reforms have certain general characteristics that entail that any evaluation of them must address and balance different and often conflicting considerations, which manifest themselves in various ways as paradoxes, dilemmas, limits and trade-offs connected with the organization and use of the evaluation (Hood & Peters, 2004; Pollitt & Bouckaert, 2011b). These are clearly evident in our empirical analyses of the evaluation of the NAV reform, and we argue theoretically that these challenges are of a more general nature.

On the basis of the analyses, we conclude that there are eight general points of attention connected with the organization of evaluations of large-scale public sector reforms, which decision makers should pay particular attention to.

Below, we start with a brief section on theory, in which we delimit the phenomenon of evaluating large-scale public sector reforms, give an account of the very limited earlier research within the area, and explain the key concepts of the analysis. After briefly explaining the methods used in the meta-evaluation, we present the results in the form of eight points of attention. The article ends with a brief conclusion.

Theory and previous research on the evaluation of reforms

“... while the fields of public-sector reform and of evaluation have each brought forth a huge body of literature and research, these two realms have been largely treated as separate entities” (Wollmann, 2003: 1).

The international research-based knowledge of evaluations of large-scale public-sector reforms is limited. The scant literature available indicates that large-scale public sector reforms are seldom evaluated (Breidahl, Gjelstrup, Hansen, & Hansen, 2017 forthcoming; Pollitt, 1995; Wollmann, 2003) and that they, in the cases where they are evaluated, are typically evaluated internally, as was the case with the Danish Local Government Reform (LGR). Although there have been numerous large-scale and minor public sector reforms in recent decades (Pollitt & Bouckaert, 2011a; Wollmann, 2003), actual research-based evaluations organized by a research council have been absent; thus, it is a rare case we are analysing here.

Characteristics of large-scale public sector reforms

We delimit the term *large-scale public sector reforms* to reforms that have an impact on substantial parts of society and cuts across governmental levels and policy sectors. Thus they are multi-level, multi-site and multi-sector reforms and, as in the case of the NAV-reform in Norway, usually mandated by a National Parliament or, in the case of federal systems, regional Parliament.

Such reforms are characterized by affecting many people, by having several often difficult-to-reconcile objectives, by developing and changing over time, and by being irreversible in the sense that they cannot or can only with difficulty be reversed back to the starting point. The large-scale public sector reforms that have been carried out in recent decades in a large number of countries have involved paradoxes, trade-offs

and dilemmas (Hood & Peters, 2004; Pollitt & Bouckaert, 2011b), the exact significance of which has varied and depended on the context.

Many reforms are decided upon or developed in a border zone between politics and administration, but also between institutions within and outside the state. Within individual states, decisions of political significance are made, which establish the framework for politicians' possibilities to act and are not affected by elected politicians.

The objectives are often manifold, and it is in many cases unclear what is being described as the actual objectives. This may be the result of previous political compromises having focused on the formulation of objectives (March & Olsen, 1983). How the reform should be presented with regard to objectives and intent has often been more controversial than the actual content of the reform (Baier, March, & Saetren, 1986). The official objectives thus often imperfectly reflect the ambitions behind the reform.

Since political-administrative reforms are complex and affect many people and policy areas, the effects are usually difficult to predict, and the most important effects can easily be beyond the official objective and target areas (Margetts, 6, & Hood, 2010; Wollmann, 2003).

Large-scale public sector reforms are often also characterized by being irreversible. The reform may be adjusted, but it is impossible to go back to doing things the way they were done before the reform without paying a high political and economic price.

The complexity and scale of the reform create some challenges for the evaluator. First, the actual nature of the reform is often modified during the implementation process and the object of the evaluation itself can therefore be difficult to pin down. Second, the reform is only one of the many factors that affect the fulfilment of objectives. Not only do a number of other state provisions have an effect, but also many other international and domestic factors. Third is the duality of the time perspective: It takes time before the effects of a large-scale reform become apparent, but the longer the time that passes, the harder it becomes to pinpoint which changes

are due to the reform and what significance other factors have (Hansen, 2010). Fourth, the perception of the various objectives of the reform and their interrelationship is not obvious. And finally, the context in which the evaluation may be intended to be used is often different from the context in which the evaluation is carried out. In other words, “reality” has changed during the process.

These factors, and others, help underline that there are considerable limitations in the use of knowledge that evaluations of large-scale political reforms can contribute to. This is especially true of questions like “How did it go?”, “What was the result?”, “Were the objectives achieved?” etc.

Attention points, paradoxes, dilemmas, trade-offs and limits

In the following analysis section, the attention points identified in connection with the meta-evaluation of the NAV reform are analysed.

By attention points we mean critical points that the key players in the evaluation – the commissioning party, the evaluator and those who are being evaluated – must be aware of, address and make decisions on. Attention points in relation to reforms and evaluations of them often arise from a balancing act between two opposing but legitimate considerations (Margetts et al., 2010; Pollitt & Bouckaert, 2011b). It may be a case of more or less sharp *trade-offs*, where assigning greater priority to one consideration must necessarily mean giving lower priority to another. There may also be some *limits* – for example, the consideration of independent research – which must be complied with if the evaluation is to be considered research-based. There may also be full-blown *dilemmas*, which mean that the evaluation will be criticized, regardless of the results it may produce. Sometimes, however, *paradoxes* – in the sense of apparent contradictions – turn out to be precisely that: apparent. If they are dealt with in a pragmatic and solution-oriented way, taking into account the specific context in which the evaluation is being conducted, they can in practice prove to be false contradictions.

Data generation and methods of analysis

Data generation took place through analyses of documents and interviews with key players in the NAV evaluation in the period from February to June 2014.

In order to identify the stakeholders' perceptions and assessments of the organization, processes, output and application of the evaluation, twelve qualitative interviews were conducted with key players in the evaluation as representatives of the main stakeholders of the evaluation: the Norwegian Ministry of Labour and Social Affairs, the Norwegian Labour and Welfare Administration (NAV), KS (Kommunernes Sentralforbund), the interest group and member authority for municipalities in Norway, the Research Council of Norway, the steering committee and a number of selected researchers. These interviews were transcribed and subsequently coded. The resources of the metaevaluation in terms of time and funding did not permit us to include all relevant stakeholders and especially the decision to exclude the politicians from the analysis may imply that some important problems related to the politics of evaluation has been omitted.

With a view to identifying the output, as available in the publications, an analysis was made of the publications from the NAV evaluation that was published up until the end of May 2014 (a total of 87 publications). These have been read and coded in order to identify the content that is relevant for the meta-evaluation, concentrating on the focus of the publications, the types of questions and the methods used.

The results of the empirical studies were analysed and discussed at the meetings of the expert panel in March, April, May and June 2014. On this background, the panel of experts developed a number of general attention points, which were presented at a concluding data generation seminar for the main stakeholders of the evaluations in mid-June 2014.

On the basis of feedback from the seminar and further discussions by the panel of experts, the final report was submitted early July 2014 with the revised eight attention points (Breidahl et al., 2014). What we present below is a further theorized version of the eight attention points found in our empirical analysis. In our analysis we attempt to show how the eight attention points represent some fairly universal trade-offs and dilemmas likely to be present in most if not all evaluations of large-scale public sector

reforms. While we believe that the eight attention points are fairly universal, we do not claim that the eight points catch all important problems related to organizing such evaluations, and especially the omission of the politicians from the analysis may have created a bias towards more administrative issues.

Eight attention points when evaluating large-scale public sector reforms

Attention point 1: What advantages and disadvantages are there in connection with an external research-based evaluation compared to alternative models?

The Research Council of Norway organizes the evaluation of the NAV reform and several other political-administrative reforms in Norway. Even though there are few comparable studies of the organization of the task of evaluation (Pollitt, 1995; Wollmann, 2003), these studies indicate that this Norwegian practice is remarkable, and it is therefore worthwhile to consider its advantages and disadvantages.

The organization of an evaluation concerns a key theme in evaluation research (Dahler-Larsen & Krogstrup, 2001; Vedung, 1997), which has explored in particular the advantages and disadvantages of internal and external evaluations, respectively, in relation to the objectives of the evaluation. From an ideal-typical perspective, a distinction is made here between the objectives of accountability (control), improvement, and basic knowledge development, which in principle call for different forms of organization. Vedung (1997: 117-20) thus argues that the objectives of accountability and basic knowledge development are best achieved through a form of external evaluation, while improvement is best ensured through the organization of an internal evaluation.

In the Norwegian evaluation, all three objectives were part of the evaluation that was commissioned (Breidahl et al., 2014). The meta-evaluation of the NAV reform indicates support for Vedung's thesis in the sense that the evaluation of the NAV reform was most successful with regard to general knowledge development and to a certain extent, analyses of the effects of the reform. In contrast, there was little continuous feedback concerning improvements and adjustments in connection with the implementation of the reform.

The choice of model for how the evaluation should be organized can be discussed from several perspectives. It is not always self-evident that a specific “model” will create a better basis for the evaluation to, for example, contribute to long-term knowledge development than other models. However our empirical findings and theoretical arguments indicate that the choice of model has implications for the problems evaluators face. If a research-based external evaluation model is chosen which also is intended to generate feedback to the on-going adaptation of the reform, special attention is needed for this purpose in the evaluation process. If an internal evaluation model is chosen, which is intended to also generate trustworthy documentation and long-term knowledge, special attention is needed for this purpose in the evaluation process.

Another discussion concerns the impartiality of the external evaluator. At times, the conclusions of external evaluations conducted by consultancy firms are criticized for being “made to order” (Wildavsky, 1996). A management team will often have an interest in controlling the results of a given external evaluation, but the very suspicion of such a manipulation having taken place can undermine the credibility of the results of an evaluation. Here, the organization in the research council has the distinct advantage that such a suspicion is eliminated or at least minimized. This requires, of course, that respect for freedom of research is mandatory and that the research council is a guarantor for this. It should also be added that the legitimacy that must be presumed to be related to a research council evaluation is not unique and can also be delivered by other institutions.

The above analysis of different models for the organization of evaluations may appear somewhat superficial. However, it will hopefully give rise to more precise analyses of advantages and disadvantages of different models than have been possible within the framework of this meta-evaluation. Norway has from an international perspective chosen an unusual and noteworthy model by using the research council as the organizer. The model has certainly been effective in relation to those objectives of the evaluation, which emphasizes contributing to long-term knowledge development.

Attention point 2: What actors should be included in the formulation of the evaluation task for the evaluation of large-scale public sector reforms?

When the commissioning party formulates the evaluation task of large-scale public sector reforms, it is an advantage to involve different actors representing expertise concerning various aspects of the reform and its evaluation. This raises the fundamental question of which actors should/must formulate the task and thereby also who should have the possibility of influencing the design of the actual evaluation. This may concern in part, an involvement of researchers and experts, and in part, an involvement of the stakeholders of the evaluation who are included in the steering committee.

The involvement of researchers and experts can have the advantage that they can provide advice and guidance, which can contribute to a clarification of the task. The challenge that may arise, however, is that in small countries like the Scandinavian countries, but perhaps also in larger countries, due to relatively small research environments within certain areas, it may be the same researchers who subsequently choose to apply to participate in the evaluation. They will therefore, all things being equal, have a competitive advantage compared to other evaluation consortia. In our view, it is a question of balancing the consideration of a satisfactory formulation of the task and the regard for an equal and fair tender process. We are dealing with a trade-off between two legitimate considerations, and the way they are dealt with in the specific situation must be a managerial decision. However, it is not appropriate if problems of impartiality prevent the best possible expertise from being drawn on in the formulation of the evaluation task. An increased use of international experts may be part of the solution.

The inclusion of the evaluation stakeholders who are part of the steering committee in the actual formulation of the evaluation task can contribute to a sense of ownership and a high degree of involvement in the evaluation on the part of the steering committee. The inclusion of the stakeholders in the formulation of the evaluation task also provides an opportunity to clarify stakeholders' expectations to and understanding of what an evaluation is and what it demands of them in terms of resources and involvement.

One possibility, which concerns the inclusion of potential actors on a more general level, is inviting all potential evaluators to an open seminar, where the formulation of the evaluation task is discussed. To the extent that national experts, who may also bid for the task, might be selected for inclusion, this bias should be taken into account in the organization of the tender process. There should be enough time for consortia that have not been involved to be able to formulate a competitive bid and the tender documents should be designed in a way that facilitates this work.

Attention point 3: Which members should be included in the steering committee?

In evaluating large-scale public sector reforms there are important considerations that indicate that the most essential stakeholders should be represented in the steering committee, as was also the case in the evaluation of the NAV reform. This applies to representatives of the commissioning party, the organizations that will be implementing the reform, and the evaluators (researchers). In addition, relevant user organizations may also be represented in the steering committee. A composition like this can contribute to promoting knowledge sharing and the use of results and to ensuring that the committee is a forum where the most significant differences of opinion and practical problems concerning the evaluation can be dealt with and resolved.

It has been argued that the commissioning party should not be part of the steering committee, since this may represent a breach of the arm's length principle. In the same way it could be argued that other stakeholders, such as the organizations responsible for implementation, should not be part of the steering committee. It can be problematic if the legitimacy that is connected with autonomy and a high quality of research can be disputed.

The balance between on the one hand, the commissioning party's and other groups' legitimate interests and on the other hand, the consideration of freedom of research, can often be achieved through the procedures that are established for cooperation between the steering committee and the researchers.

Here it is again a balancing act, which depends on the characteristics of the specific reform that is to be evaluated. In the case of a strongly politicized reform, where there is a high degree of societal distrust in the role of the commissioning party, a formalization of the arm's length principle through the composition of the steering committee may be considered. The commissioning party in these cases might, however, have observer status.

Attention point 4: What principles should characterize the relationship between the steering committee and researchers?

In a research-based evaluation organized by the Research Council of Norway, great importance will naturally often be attached to long-term knowledge development, while the continuous instrumental use of preliminary results, which could be useful in connection with an adaptation and implementation of the reform, will often require particular attention. In the classical trade-off between “quick-and-dirty” or “slow-and-clean” analysis, those responsible for implementing the reform will need the former, while the basic inclination of researchers will be to stick to the latter.

Here, the steering committee, together with the project manager and research coordinators, play an essential and legitimate role as a bridge builder and intermediary between research and the implementation process. An active steering committee can furthermore help ensure that resources and competences that can be useful for the evaluation are made available. This can often have a very positive effect on the evaluation process and the quality of the results of the evaluation.

Care must nonetheless be taken to ensure that this role as a bridge builder and intermediary does not break with the arm's length principle and the consideration of high quality research. There are certain limits to intervention that must be observed to ensure that the qualities that characterize research-based evaluations are maintained and the legitimacy of the evaluation is protected.

It is reasonable that the commissioning party and the members of the steering committee are able to access information about analyses and conclusions of the evaluation before they are published. In this way, the commissioning party and the

organizations responsible for implementation can keep abreast of the situation and quickly draw on relevant results during the continuous implementation of the reform.

The commissioning party and the organizations responsible for implementation will also in some cases have factual knowledge that can contribute to improving the quality of publications made in connection with the evaluation.

However, this type of intervention must never be more than helpful suggestions. The impartiality of the researchers must never be questionable. This applies in general and is especially important in relation to evaluations of large-scale public sector reforms. With these reforms, there will often be marked interest in influencing the evaluation in certain directions. There must never be any doubt about the integrity of the research and the impartiality of the commissioning party and other players' possible interest in concealing problems in connection with the implementation process.

Attention point 5: How can the numerous and difficult-to-reconcile objectives of the evaluation be dealt with?

The choice of the research council as the organizer of the evaluation means that the commissioning party – consciously or unconsciously – has decided to prioritize long-term knowledge development as an essential element of the evaluation task. By protecting and prioritizing the research council's long and well-developed tradition of ensuring high quality, the aim has (presumably) been to ensure the quality and legitimacy of the evaluation in that the research should contribute to an arm's length principle stipulated by the commissioning party. A sub-objective has also been to contribute with the continuous applicability of the results of the evaluation in connection with the implementation.

The challenges that these different intended uses give rise to should have been subject to more extensive discussion even before the choice of the form of organization was made. There is tension if not contradiction between long-term knowledge generation and providing short-term advice for policy implementation. Part of the tension is captured in the following quote from the French sociologist and anthropologist Marcel Mauss: "In scientific matters, it is impossible to proceed too slowly; while in

practical matters one cannot wait” (Mauss quoted in Rosanvallon, 2012, p. 57)
Although the quote pushes the issue to the extreme, there is a tendency for long-term knowledge development to place other and in part conflicting demands on the evaluation than those that arise from continuous feedback on the implementation process. The dynamics of scientific research processes, both in terms of timing and reference structure (e.g. the theoretical issues considered important), is very different from the dynamics of an implementation process with a need for swift feedback concerning challenges that may be irrelevant from a scientific point of view.

Continual use of the results of the evaluation can take place on several levels, from taking a more general position, over the actual allocation of tasks and responsibilities between various players, to changes in how local units function. Those parts of the evaluation that are carried out with an eye to continuous feedback must therefore be directed at these different levels. Such an evaluation requires a close relationship between those evaluating and the various players who will be using the results of the evaluation (Hansen, 2005); this may possibly challenge autonomy and the arm’s length principle. Some evaluation researchers who emphasize direct use, for example, Patton (Patton, 1997), have therefore underlined that instead of distance and arm’s length, evaluations require nearness to those who will need to use the evaluation continuously. Furthermore, the view is put forward that evaluations with such instrumental objectives must mean that an evaluator must be involved in the firms that are to be evaluated.

Here we are dealing with a paradox – in the sense of an apparent contradiction – when the objective is both continuous feedback and long-term knowledge development. Whether it really is a case of a genuine conflict between the requirement of autonomy in long-term knowledge development and the demands of continuous useful feedback is subject to different opinions. Yet such different objectives for use do give rise to a number of challenges for the evaluator, which must be taken into account. We understand that it may be advisable from a tactical perspective to insist in the planning phase that different objectives be included in the evaluation. One way to cope with the challenge may possibly be to organize different research projects for the different purposes within the overall research design. Another may be to assign responsibilities to the different purposes to specific evaluators within each research

project. However if the challenge is not attended to it is likely to result in a negative in the outcome of the evaluation.

The evaluation of the NAV reform provides evidence of the problems that arise when evaluations with several intended uses are to be conducted. Although it is possible design an evaluation that is aimed at long-term knowledge development in such a way that it also can contribute to continuous instrumental use, *this* form of use is unlikely to be the key strength of a research agency based evaluation. Decisions on the organization of future evaluations should therefore also take into account considerations of the extent to which different types of organizations are conducive to different intended uses.

Attention point 6: How is flexibility incorporated in the evaluation process?

In the theory section, we described certain characteristics of large-scale public sector reforms. Among other things, it was pointed out that the real nature of the reform to a high degree is developed in the implementation process itself. This has become very clear in the case of the NAV reform. Whether it is the intended reform or the actual reform that is being evaluated will therefore depend on how the evaluation is decided upon and designed. It is important to strive to evaluate the reform that actually has been carried out.

Various ideas about how to do so can be found in the empirical study in the meta-evaluation of the evaluation of the NAV reform. The most important are first, to maintain a ‘wait and see’ attitude during the first phase of the evaluation in order to observe how the reform develops in practice, and second, and related to this, to reserve a pool of funds for unforeseen analyses that prove to be particularly important during the evaluation. In this way the steering committee, in cooperation with the project and module managers are given a more significant task than in traditional research projects of developing the evaluation design while reflecting on how the reform is developing. The disadvantage can be a continuous “political tug-of-war” for these funds.

In any case, the changeability of the reform itself during the implementation process must be seen as a key question in the evaluations. This aspect was not taken adequately into account in the decision-making and design of the NAV evaluation. It should therefore be discussed in more detail in connection with future evaluations.

Attention point 7: How ambitious should one be about integrating different research modules in an evaluation?

In the empirical study in the meta-evaluation of the evaluation of the NAV reform, several researchers raised the question of how ambitious one should be with regard to coordinating and integrating different research modules in a larger joint research project, and whether, alternatively, one should to a higher degree work with more loosely connected project coordination or several independent research projects. The argument against ambitious integration is that much time and many resources can be wasted on trying to integrate different research traditions that are very difficult if not impossible to bridge. The argument in favour of ambitious integration is that different research projects, whose connections have been thoroughly thought through, can strengthen the credibility of the overall research design.

This is an important attention point, which depends on an assessment of the specific evaluation task. The empirical analyses of the meta-evaluation and the final conference of the NAV evaluation suggest, however, that considerable progress has been made with the coordination strategy, despite numerous frustrations along the way. The process evaluations thus provide credible suggestions for the processes that can explain the results of the effect evaluations. Furthermore, various recent trends within evaluation theory (Chen, 2014; Hansen & Vedung, 2010; Pawson, 2013) indicate that mapping of the links between cause and effect in process evaluations are often decisive for being able to determine whether effects that are identified in measurements of effect are the result of the reform or of other circumstances – and even more importantly, what characteristics of the reform and its context have brought about the effects.

Attention point 8: Should the possibilities for evaluating reform impact be incorporated in the reform decision and the implementation of reforms?

The challenges encountered in the NAV evaluation in terms of identifying effects bring into focus the discussion of the extent to which evaluation should be taken into account in connection with the organization of large-scale public sector reforms. Should reforms ideally be organized in such a way that “proper” measurements of effect can be carried out, organized as quasi-experiments with (ideally) randomized control groups and before and after measurements? This is an old idea from the 1960s and practice since then indicates that it is seldom – if ever – possible – in the case of large-scale public sector reforms.

To begin with, there should be a consensus of opinion today, also between the different sides of the evidence debate, on the fact that the possibilities to make statements with certainty about the effects of large-scale public sector reforms are very limited. The different institutes (e.g. Cochrane and Campbell) that work with these questions have shown that most evaluations by far, including those carried out by well-established researchers, do not contribute with this sort of knowledge. The other side of the debate has, among other things, discussed the very question of what type of knowledge can be considered evident and what should fundamentally be understood by evidence. We therefore consider these points of view in principle as unrealistic, not just with regard to how politics comes into being but also with regard to the possibilities that the production of knowledge creates.

This point of view is also problematic from a democratic perspective since it implies that politicians’ demands are subordinate to the production of knowledge in society and the demands that are put forth by evaluators. This raises the question of what role the measurements of effect should have and how they should be organized in connection with large-scale public sector reforms. As in the case of the NAV-reform, sometimes it may be possible to organize large-scale effect measurements of important aspects of the reform as a part of the evaluation design (Andreassen and Aars 2015). Sometimes it may be possible in retrospect to reconstruct the effects of reform efforts with reasonable validity based on register and a combination of secondary data as in a recent analysis of the impact the New Public Management reforms in the UK (Hood & Dixon, 2015). This is however not a question that can be answered in general terms but it is an important point of attention when organizing evaluations of large-scale public sector reforms.

Conclusion

The eight attention points of the meta-evaluation discussed above have been found through an analysis of a single and rare case – the evaluation of the NAV reform in Norway – and we have argued that they are likely to be of relevance to most if not all evaluations of large-scale public sector reforms. The eight attention points are listed in table 1 below.

*** Around here table 1***

We have worded the eight points as questions, which decision makers involved in the evaluation of large-scale public sector reforms should take into consideration, and we have attempted to show the typical issues of balancing and trade-offs that characterize each attention point.

The eight attention points are likely to have general relevance and are essential to consider and to take into account when organizing future evaluations of large-scale public sector reforms and many of them are likely to have more general relevance to the organization of other types of evaluations.

References

- Andreassen, T. A., & Aars, J. (2015). *Den store reformen. Da NAV ble til*. Oslo: Universitetsforlaget.
- Baier, V. E., March, J. G., & Saetren, H. (1986). Implementation and ambiguity. *Scandinavian Journal of Management Studies*, 2(3-4), 197-212.
- Breidahl, K. N., Furubo, J.-E., Halvorsen, K., & Hansen, M. B. (2014). *Opmærksomhedspunkter ved evaluering af politiske reformer. Rapport fra metaevalueringen af NAV-reformen i Norge*. Retrieved from Oslo:
- Breidahl, K. N., Gjelstrup, G., Hansen, M. B., & Hansen, H. F. (2017 forthcoming). Evaluation of Large-scale Public Sector Reforms: A Comparative Analysis. *American Journal of Evaluation*.
- Chen, H. T. (2014). *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective* (2 ed.). Thousand Oaks, California: SAGE Publications.
- Dahler-Larsen, P., & Krogstrup, H. K. (2001). *Tendenser i evaluering*. Odense: Odense Universitetsforlag.
- Hansen, M. B. (2005). *Anvendelse af analyser i beslutningsprocesser*. Retrieved from København:
- Hansen, M. B. (2010). Marketization and Economic Performance. Competitive Tendering in the Social Sector. *Public Management Review*, 12(2), 255-274.

- Hansen, M. B., & Vedung, E. (2010). Theory-Based Stakeholder Evaluation. *American Journal of Evaluation*, 31(3), 295-313. doi:10.1177/1098214010366174
- Hood, C., & Dixon, R. (2015). *A Government That Worked Better and Cost Less?: Evaluating Three Decades of Reform and Change in UK Central Government*. Oxford: Oxford University Press.
- Hood, C., & Peters, G. (2004). The middle aging of new public management: Into the age of paradox? *Journal of Public Administration Research and Theory*, 14(3), 267-282.
- March, J. G., & Olsen, J. P. (1983). Organizing Political Life - What Administrative Reorganization Tells us About Government. *American Political Science Review*, 77(2), 281-296.
- Margetts, H., 6, P., & Hood, C. (Eds.). (2010). *Paradoxes of modernization: Unintended Consequences of Public Policy Reform*. Oxford: Oxford University Press.
- Patton, M. Q. (1997). *Utilization-focused Evaluation*. *The New Century Text*. Thousand Oaks: SAGE.
- Pawson, R. (2013). *The science of evaluation: a realist manifesto*. London: SAGE publications.
- Pollitt, C. (1995). Justification by Works or by Faith? Evaluating the New Public Management. *Evaluation*, 1:2, 133-154.
- Pollitt, C., & Bouckaert, G. (2011a). *Public Management Reform. Comparative Analysis of New Public Management, Governance and the Neo-Weberian State* (Third ed.). Oxford: Oxford University Press.
- Pollitt, C., & Bouckaert, G. (2011b). Trade-Offs, balances, limits, dilemmas, contradictions and paradoxes. In C. Pollitt & G. Bouckaert (Eds.), *Public Management Reform. Comparative Analysis of New Public Management, Governance and the Neo-Weberian State*. Oxford: Oxford University Press. (pp. 182-205).
- Rosanvallon, P. (2012). *Democracy Past and Future: Selected Essays*: Columbia University Press.
- Stufflebeam, D. L. (2001). The metaevaluation imperative. *American Journal of Evaluation*, 22(2), 183-209. doi:10.1177/109821400102200204
- Vedung, E. (1997). *Public Policy and Program Evaluation*. New Brunswick: Transaction Publishers.
- Wildavsky, A. (1996). *Speaking Truth to Power. The Art and Craft of Policy Analysis*. New Brunswick: Transaction Publishers.
- Wollmann, H. (Ed.) (2003). *Evaluation in public-sector reform: concepts and practice in international perspective*. Cheltenham UK: Edward Elgar Publishing.

End note

¹ A meta-evaluation should not be confused with meta-analyses. A meta-evaluation is simply an analysis of the value of one or more evaluations (Stufflebeam, 2001) that are intended to be used in future decision-making processes.

1	<i>What advantages and disadvantages are there in connection with an external research-based evaluation compared to alternative models?</i>
2	<i>What actors should be included in the formulation of the evaluation task for the evaluation of large-scale public sector reforms?</i>
3	<i>Which members should be included in the steering committee?</i>
4	<i>What principles should characterize the relationship between the steering committee and researchers?</i>
5	<i>How can the numerous and difficult-to-reconcile objectives of the evaluation be dealt with?</i>
6	<i>How is flexibility incorporated in the evaluation process?</i>
7	<i>How ambitious should one be about integrating different research modules in an evaluation?</i>
8	<i>Should the possibilities for evaluating reform impact be incorporated in the reform decision and the implementation of reforms?</i>