



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

Evaluation of Large-scale Public Sector Reforms

A Comparative Analysis

Breidahl, Karen Nielsen; Gjelstrup, Gunnar; Hansen, Hanne Foss; Hansen, Morten Balle

Published in:
American Journal of Evaluation

DOI (link to publication from Publisher):
[10.1177/1098214016660612](https://doi.org/10.1177/1098214016660612)

Publication date:
2017

Document Version
Accepted author manuscript, peer reviewed version

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Breidahl, K. N., Gjelstrup, G., Hansen, H. F., & Hansen, M. B. (2017). Evaluation of Large-scale Public Sector Reforms: A Comparative Analysis. *American Journal of Evaluation*, 38(2), 226-245.
<https://doi.org/10.1177/1098214016660612>

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.

Evaluation of Large-Scale Public-Sector Reforms: A Comparative Analysis

American Journal of Evaluation
1-20

© The Author(s) 2016

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/1098214016660612

aje.sagepub.com



**Karen N. Breidahl¹, Gunnar Gjelstrup²,
Hanne Foss Hansen², and Morten Balle Hansen¹**

Abstract

Research on the evaluation of large-scale public-sector reforms is rare. This article sets out to fill that gap in the evaluation literature and argues that it is of vital importance since the impact of such reforms is considerable and they change the context in which evaluations of other and more delimited policy areas take place. In our analysis, we apply four governance perspectives (rational-instrumental perspective, rational interest-based perspective, institutional-cultural perspective, and chaos perspective) in a comparative analysis of the evaluations of two large-scale public-sector reforms in Denmark and Norway. We compare the evaluation process (focus and purpose), the evaluators, and the organization of the evaluation, as well as the utilization of the evaluation results. The analysis uncovers several significant findings including how the initial organization of the evaluation shows strong impact on the utilization of the evaluation and how evaluators can approach the challenges of evaluating large-scale reforms.

Keywords

evaluation use, meta-evaluation, multilevel evaluation, governance, reform

The number of large-scale public-sector reforms aimed at changing political and administrative structures and processes have increased in many European and Anglosphere countries since the 1980s (Pollitt & Bouckaert, 2011a). However, in spite of the significance and impacts of these reforms, a country comparison in 2003 showed that evaluation researchers had underexplored these reforms:

Public-sector reform and evaluation have been closely interlinked almost like Siamese twins throughout the past 30 years or so. Yet an inspection of the available literature on public-sector reforms and evaluation reveals a glaring discrepancy: while the fields of public-sector reform and of evaluation have

¹ Department of Political Science, Aalborg University, Aalborg, Denmark

² Department of Political Science, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:

Morten Balle Hansen, Department of Political Science, Center for Organization, Management and Administration (COMA), Aalborg University, Fibigerstræde 1, 9220 Aalborg, Denmark.

Email: mbh@dps.aau.dk

each brought forth a huge body of literature and research, these two realms have been largely treated as separate entities. (Wollmann, 2003b, p. 1)

The interlinkage between public-sector reform and evaluation can be analyzed from two perspectives. One perspective asks to what degree the diffusion of evaluation has been part of and contributed to the diffusion of ideas about public administration, and another perspective asks to what degree, why, and how the evaluations of public-sector reforms have been conducted. This article is concerned with the second perspective. Focus is on large-scale public-sector reform which is defined as reforms cutting across governmental levels and policy fields. Scholars within the field of comparative public administration have also in recent years concluded that the evaluation of public-sector reforms has largely been absent (Kuhlmann & Wollmann, 2011) and remains a rarity (Kuhlmann & Wollmann, 2014) especially in relation to reforms aimed at changing subnational (regional and local) levels. As we will account for thoroughly in the next section, our own updated review of international literature on research on the evaluation of public-sector reforms in the period 1995–2015 does not challenge this conclusion: Research on the evaluation of large-scale public-sector reforms is for sure an exception and compared to the countless substantive sector reforms in for instance the health-care sector and the education system large-scale public-sector reforms are a rare evaluand (Pollitt, 1995, p. 135; Pollitt & Bouckaert, 2011b; Wollmann, 2003b).¹

This reflects how most evaluation literature (including the North American) for several years mainly have been concerned with specific program and intervention evaluations rather than research on to what degree, why, and how the evaluations of public-sector reforms have been conducted.

In this article, we argue that not only is knowing how to evaluate large-scale public-sector reforms of critical importance; so too is knowing how to conduct research on and evaluate the nature of these evaluations, their utilization, and the broader role they play in public-sector reforms. This is a challenging endeavor: Evaluations of large-scale public-sector reforms involve a very high degree of complexity, as this type of reform cuts across governmental levels and policy sectors. Large-scale public-sector reforms are further as also many sectorial reforms surrounded by much political interest and their (often) ambiguous aims are based on political compromises and involve a heterogeneous group of actors. Furthermore, their target groups are complex and they imply complex issues that are implemented over a long time period, whereby the contexts of evaluations are always in flux. Together, these characteristics make it difficult both to evaluate these reforms and to evaluate the evaluations of these reforms (Andreassen & Aars, 2015; M. B. Hansen, Breidahl, Furubo, & Halvorsen, in press; M. B. Hansen, Breidahl, Halvorsen, & Furubo, 2015; Pollitt, 1995).

Nevertheless, it is vital to better understand the dynamic and complex nature of large-scale public-sector reforms and evaluations of them. Due to their large-scale nature, the impact of these reforms is considerable and they also change the structural boundaries and context in which the evaluations of other and more defined policy areas within the public sector take place (Pollitt, 1995). It is therefore important to better grasp the complexity of these evaluations and their consequences. Furthermore, from a more instrumental point of view, a better understanding of these evaluations is important for successful implementation and for the sustainability of existing reforms.

This article intends to fill this research gap. In order to do so, it systematically compares and examines recent evaluations of two large-scale public-sector reforms in two Nordic countries: the reform of the labor and welfare administration in Norway (NAV reform) and the local government reform in Denmark (LGR reform). The reforms, which were carried out in the 2000s, represent the most important public-sector reforms in the two countries in recent decades and have a number of characteristics in common: They were adopted centrally by Parliament (Storting/Folketing), they affected large parts of the public sector and the population, they had several and partly incompatible objectives, and they developed and changed over time. Furthermore, the two reforms were characterized by a focus on formal structural change, as the main policy instrument and included the

state, regional, and local levels of public administration. Finally, and crucial in this context, both reforms have been evaluated. Systematically comparing and examining these two cases allows us to contribute to a neglected research field and shed light on the complexity of the evaluation of large-scale public-sector reforms and hopefully on how evaluators can approach the substantial challenges of evaluating this type of reform.

Despite the similarities of the reforms, the evaluations were organized, implemented, and utilized in very different ways and with very different impacts. While the evaluation of the NAV reform in Norway was organized as a research-based external evaluation, the evaluation of the Danish municipal reform was organized as an internal evaluation, with evaluation responsibilities resting with key stakeholders. Our analysis shows that the way these evaluations were organized had significant impact on the output as well as the outcomes of the evaluations, including how they were utilized. Hence, while the evaluation of the NAV reform was used mainly for knowledge creation and building a democratic debate, the evaluation of the local government reform was used instrumentally to adjust the division of labor between levels of government in selected areas of activity.

These notable differences between the two evaluations raise three research questions, which we will explore in the empirical analysis: (1) How were the two evaluations initiated and carried out? (2) How were they utilized? and (3) How can we theoretically interpret the main differences and similarities between the two evaluations? While the first two questions are answered through a comparative analysis of the core elements of the reforms and the process, organization, and utilization of the two evaluations, the third question is answered through a theoretically informed analysis based on four governance theory positions: a rational-instrumental perspective, a political interest-based perspective, an institutional-cultural perspective, and a chaos perspective. Applying these different theoretical perspectives allows us to better grasp the complexity of the two evaluations and thereby contribute to the existing evaluation literature, where these insights are more or less absent. Each of the perspectives also raises several questions that we will deal with in the final section of the article.

The article unfolds as follows: We start by clarifying the concept of large-scale public-sector reforms and present and discuss the findings of the international literature review. Moreover, we outline the research design and methodology and the theoretical framework. Two analytical sections follow: The first compares the two evaluations in terms of the (1) focus of the evaluation (the evaluand) and the purpose of the evaluation, (2) organization of the evaluation, and (3) results and utilization of the evaluations. The second examines and discusses how the nature of the two evaluations and the differences between them can be interpreted according to the four governance theory perspectives. The last section presents the conclusions and reflects on the implications of these for handling the challenges of evaluating large-scale public-sector reforms.

Evaluating Large-Scale Public-Sector Reforms: Conceptualizations and Existing Knowledge

As outlined in the Introduction section, a systematic international literature review in the period 1995–2015 has been conducted in order to outline existing knowledge on research on the evaluation of large-scale public-sector reforms. However, before presenting the findings of the review, the concept of large-scale public-sector reforms needs to be clarified.

The literature on public-sector reforms is characterized by conceptual ambiguity. Many concepts are used, such as administrative reforms, public management reforms (new as well as post-new), and (new) public governance reforms. We use the concept of public-sector reforms as a generic term covering several types of reforms, all of which refer to *formal changes in political and administrative processes and structures in institutions of governance*. Thus while such reforms may also include “substantive policies” targeted at a concrete policy goal or policy output (crime prevention, reduction of unemployment, etc.), their prime focus is on the processes and the structures of the

public sector. Public-sector reforms are complex and follow several steps. The aim is to change the political and administrative institutions, but these institutional changes are also intended to bring further results "... whether it be that the operational process ('performance') of public administration or that the (final) 'output' and 'product' of the administrative operation is improved'" (Wollmann, 2003b, p. 5). Or as Pollitt and Bouckaert (2011a, p. 2) frame it to change "... the structures and processes of public sector organizations with the objective of getting them (in some sense) to run better." Public-sector reforms are therefore not defined by an exclusion of substantive policies, but the primary focus is on change in governance institutions.

We are concerned not with all types of public-sector reforms but more precisely with large-scale public-sector reforms. Large-scale reforms are reforms of considerable size. In line with H. F. Hansen (2005b), size here refers to reforms cutting across governmental levels and policy fields. Large-scale public-sector reforms are thus defined as multilevel, multisite, and cross-sectional reforms aiming at changing political and administrative institutions.

As mentioned in the Introduction section, only one country comparative study on the evaluation of large-scale public-sector reforms (mainly New Public Management reforms) has been carried out. One of the conclusions drawn is that these reforms are most often organized as internal evaluations rather than external evaluations conducted "... by an agency or actor outside of and different from the operating unit," which is a rarity in all countries (Wollmann, 2003b, pp. 6, 250). Similarly, an article comparing evaluations in and of large-scale public-sector reforms in Denmark from 1982 to 2005 to the Scandinavian countries finds that "Evaluations of reform activities have been partial, seldom overall and mostly internal" (H. F. Hansen, 2005, p. 344).

In order to update and substantiate these findings, a systematic literature review on research on the evaluation of large-scale public-sector reforms has been conducted of nine central international journals on evaluation, public administration, and public policy: (1) *American Journal of Evaluation*, (2) *Canadian Journal of Program Evaluation*, (3) *Evaluation*, (4) *Evaluation Review*, (5) *Evaluation and Program Planning*, (6) *Public Administration*, (7) *Public Administration Review*, (8) *Journal of European Public Policy*, and (9) *Journal of Public Policy* in the period from 1995 to 2015. The review has not included articles on specific evaluation of large-scale public-sector reforms but has been focusing on research on the evaluation of large-scale public-sector reforms and questions like, for example, to what degree, why, and how the evaluations of public-sector reforms have been conducted. All possible knowledge of this subject has not been exhausted, but the selected evaluation and public policy journals represent the vast majority of contributions within the area of evaluations of public-sector reforms.

Due to the definitional challenges of defining the concept of "large-scale public-sector reforms" and conceptual ambiguity, it was necessary to apply a rather broad search strategy. Hence, all articles combining the two terms "reform" and "evaluation" were included in order to make sure that relevant articles were not excluded from the search results. Consequently, all articles combining evaluation and a broad group of reforms ("administrative," "management," "governance," "large-scale," "welfare state," "multilevel," etc.) was included in the first round of search results for each journal. This included up to 800 articles in some of the journals. Afterward, in the second search round, all the articles were sorted by the authors based on a qualitative assessment of whether or not they focused on the evaluations of large-scale public-sector reforms (based on our definition compared above) and whether they focused on research on this type of evaluations or specific evaluations. After this sorting process, the number of relevant articles was dramatically reduced to a select few. Hence, most of the articles from the first selection of articles were concerned with the evaluations of substantial policy sector areas such as educational reforms, welfare to work programs, and so on. The final selected group of articles included only few articles focusing on the evaluations of large-scale public-sector reforms (Van Eyk, Baum, & Blandford, 2001) and a few articles reporting on specific evaluations that have been conducted of public-sector reforms

(Kuhlmann, Bogumil, & Grohs, 2008; Pedersen & Rieper, 2008). Another article was concerned with problems related to not evaluating new public-sector reforms and why there is resistance to doing so (Broadbent & Laughlin, 1997). Finally, one article discussed whether inputs from evaluations are used in reforms of administrative and public-sector management (Thoenig, 2000).

A clear conclusion from the literature review is that existing research on evaluations of large-scale public-sector reforms (to what degree, why, and how the evaluations of public-sector reforms have been conducted) is scarce. Thus, the present analysis fills an important gap in the literature.

Research Design, Empirical Material, and Methodology

The research design is based on a comparison of two cases that are in some ways similar, but in other ways different, especially in terms of the organization and utilization of the evaluations. Thus, we consider the comparison of the two as a critical case for the study of evaluations of large-scale public-sector reforms. A deep understanding of differences and similarities can form a starting point for considerations of how to handle dilemmas in the evaluation of such reforms. This is also the reasoning behind making the comparison quite detailed. The comparative analysis of the two evaluations (the NAV evaluation and the LGR evaluation) is based on a broad range of documents (including minutes from meetings, policy documents, reports, different bills, and evaluation proposals) as well as a systematic review of all material published as part of the evaluations—about 87 reports and articles in Norway and 13 reports in Denmark. Furthermore, semistructured interviews have been conducted among a number of key figures and stakeholders in both countries. Thirteen interviews were conducted in the spring of 2014 in Norway among civil servants representing the public evaluation sponsors; evaluators (researchers) responsible for the design, execution, and findings of the evaluation; representatives from the Norwegian Research Council; members of the steering committee appointed by the research council; the national agency managers from the Norwegian Labor and Welfare Service; and a representative from the Norwegian Association of Local and Regional Authorities. As there is a large and transparent body of documentary material of the Danish evaluation, there was no need for a large number of interviews in relation to this case. Therefore, four interviews were conducted in the spring of 2013 with key participants in the evaluation process. Of these, three were members of the committee responsible for conducting the evaluation, and one was a member of one of the four subgroups responsible for the interim reports, which formed part of the input that led to the overall report. The interview guides consisted of a range of themes, including process, organization, results, outcomes, utilization, and administrative follow-up (Breidahl, Furubo, Halvorsen, & Hansen, 2014; Gjelstrup & Hansen, 2014). The interviews have been transcribed. An analysis of the data has been published previously in Danish and Norwegian (Breidahl et al., 2014; Breidahl, Gjelstrup, Hansen, & Hansen, 2015; Gjelstrup & Hansen, 2014; M. B. Hansen, Breidahl, Halvorsen, et al., 2015).

Theoretical Framework

Theoretically, the analysis draws on four governance perspectives, including two rational perspectives (one instrumental and one interest based), one institutional-cultural perspective, and a perspective that is the outright antithesis of the rational: the chaos perspective. This analytical strategy positions the analysis in line with Graham T. Allison's classic argument, better enabling us to grasp the complexity of the two evaluation processes (Allison, 1971). The analytical strategy followed is a "filling strategy" (Grøn, Hansen, & Kristiansen, 2014; Roness, 1997), sometimes also referred to as a complementary strategy. This means that the four perspectives in the framework supplement each other, as they contribute to the interpretation of the two evaluations as well as of the different phases in the two evaluation processes. Table 1 gives an overview of how organizational activities, actors, and reform implementation processes are understood in the four different perspectives.

Table 1. Four Conceptual Models for Analyzing the Evaluation of Public-Sector Reforms.

| Dimensions | Rational-Instrumental Perspective | Rational Interest-Based Perspective | Institutional-Cultural Perspective | Chaos Perspective |
|-------------------------|---|--|---|--|
| Activities are | Collective, system-based and anchored in goal-means rationality | Developed as results of negotiations | Based on norms, values and standard operating procedures (SOPs) | Emergent, since they take place in contexts of ambiguity, complexity, and changeability |
| Actors are | Compliant and loyal | Promoting narrow self-interest | Adapting to what is seen as appropriate | Characterized by shifting attention to problems and solutions, implying that connections between these become random |
| Reforms are implemented | Top down in planned and linear processes | Through ongoing negotiation processes, implying frequent adjustments and changes | Through translation processes, adapting the content to the cultural context | In emergent and unpredictable processes |

The *rational-instrumental perspective* is rational in a collective and system-based sense (Scott & Davis, 2007, chap. 2). It is based on the idea that means are selected that will match and promote system and organizational objectives. Actors are compliant and reforms are implemented in planned and linear processes.

The concept of rationality underlining the *interest-based perspective* is narrower. The interest-based perspective sees the world as consisting of groups, organizations, and fields of organization that, through negotiations and bargaining, seek to promote their own interests as far as possible (Scott & Davis, 2007). In this perspective, a reform process is considered an arena for debate and political bargaining. The negotiations may concern the content of the reform as well as its implementation.

According to the *institutional-cultural perspective*, attitudes, norms, values, and traditions govern the behavior of individuals, organizations, and fields of organization. The stakeholders are expected to follow the informal rules of the game, which evolve over time in a collective learning process related to how challenges are addressed (March & Olsen, 1989; Schein, 1985). According to this perspective, reforms are implemented through translation processes adapting the reform content to the cultural context.

Finally, the *chaos perspective* emphasizes the importance of ambiguity, complexity, and changeability for the way in which organizations and fields of organization develop. Organizations and fields of organization are understood as emergent, that is they do not change because of linear processes that can be explained by the existing organization, but instead in ways that can be explained based on their loosely connected ecological nature (March, 1999; Stacey, 2011; Weick, 2001). The chaos perspective, therefore, does not consider processes of organizational change to be irrational or, generally speaking, incremental. Rather, random factors caused by time-specific connections between the stakeholders' attention and resources, as well as their connections to solution and problem streams, influence reform processes (March & Olsen, 1986).

Table 2. Four Perspectives on Essential Dimensions of Evaluation.

| Dimensions | Rational-Instrumental Perspective | Rational Interest-Based Perspective | Institutional-Cultural Perspective | Chaos Perspective |
|------------------------------|---|---|--|---|
| Evaluation processes are | Organized according to agreements on evaluation goals | Arenas for negotiations | Culturally determined | Emergent and unpredictable |
| Evaluators play the roles of | Experts | Facilitators of negotiations between stakeholders and/or promoters of self-interest | Creators of meaning | Creators of platforms for further development processes |
| Utilization | Evaluation results are used in instrumental ways for controlling the results of reforms and deciding on adjustments | Evaluation processes as well as results are used in tactical ways as stepping-stones for stakeholders' promotion of self-interest | Evaluation processes are used to create meaning and results are used to legitimize the reforms | Both evaluation processes and results are used in unpredictable garbage can-like ways |

The theoretical framework with the four perspectives can be applied to important dimensions put forward in the evaluation literature to analyze and characterize specific evaluations. Table 2 provides an overview of how the evaluation processes, the role of evaluators, and the ways in which evaluations are utilized are understood in the different perspectives, thereby contributing to the development of evaluation theory and practice.

Applied to our focus on evaluation, the *rational-instrumental perspective* emphasizes that the purpose of the evaluation determines its design. According to this perspective, we would expect the democratically elected political leadership to agree on an evaluation goal and further, that this goal would determine the organization and design of the evaluation, which constitute the framework of the analyses, assessments, and conclusions. In this perspective, evaluators are expected to play the role of neutral experts, carrying out the evaluation and reporting the results to the decision makers and the actors responsible for the implementation of reforms.

Applied to our focus on evaluation, the *rational interest-based perspective* emphasizes that evaluation processes are considered arenas for debate. The negotiations may concern the timing, purpose, focus, and organization of the evaluation, as well as the conclusions drawn and solutions proposed during the follow-up. Evaluators act either in the roles of facilitators of negotiation or as promoters of self-interest. Both evaluation processes and results are used in tactical ways.

Applied to our focus on evaluation, the *institutional-cultural perspective* emphasizes that general public-sector values, as well as more specific evaluation values, would influence the design and organization of concrete evaluations. Evaluation processes are used by evaluators for the creation of meaning and evaluation results, for legitimizing.

Finally, applied to our focus on evaluation, the *chaos perspective* emphasizes the emergent nature of the evaluation processes. Evaluation processes are assumed to be complex and dynamic processes evolving in unpredictable ways. Evaluators may actively create platforms to further development processes. Utilization is unpredictable as both processes and results may be coupled in garbage can-like processes to agendas outside the reform and evaluation focus.

The four perspectives will be used in both the analysis and the discussion of the way the two evaluations were organized and designed, and in the broader analysis of the reform processes and the follow-up to them.

Two Large-Scale Public-Sector Reforms and Two Evaluations: A Comparative Analysis

In this section, the results of the comparative analysis are presented. First, the two reforms are briefly introduced and their core elements are compared. Then, the evaluations of the two reforms are compared on the dimensions discussed above in the theoretical framework (the evaluation process, the evaluators, and the utilization of evaluation results). However, before turning to the analysis the context will be presented.

Context

Norway and Denmark share many features. Both countries are decentralized unitary states (H. F. Hansen, 2011; M. B. Hansen, Lægneid, Pierre, & Salminen, 2012). Both are well organized and have long traditions for involving stakeholders in policy development and implementation. Both are also renowned for providing relatively high social protection and generous benefits, for their universal, predominantly tax-financed welfare state arrangements, and for their high degree of government intervention and welfare state services, which are mainly provided by public-sector institutions (Brochmann & Hagelund, 2012).

In both countries, evaluation practice was for many years embedded in separate disciplinary scientific fields (Hansson, 1998; Haug, 1998), influenced more by continental philosophy than by logical empiricism (Schwandt, 1998). But in the last 15 years, evaluation in both countries has developed into a field in its own right with a broad group of actors taking many approaches and engaging in discussions in associations and at conferences. Partly as a result of this, the evaluation cultures in both countries are characterized as mature (Jacob, Speer, & Furubo, 2015). In this context, it could be expected that evaluating large-scale public-sector reforms like the ones in focus here would be taken for granted in both countries.

Analytical Comparison of the Two Public Management Reforms

The core dimensions of the two large-scale public-sector reforms are summarized in Table 3.

The labor and welfare reform in Norway came into force on January 1, 2006 (Andreassen & Aars, 2015; Fimreite & Lægneid, 2009). The reform was comprehensive and entailed the amalgamation of three services (the employment service, the national insurance administration, and the social services). Moreover, the reform implied partnerships between the new labor and welfare administration on the one hand and local governments on the other. The overall goal was to reduce the share of persons in the workforce receiving social benefits, and the three objectives were integration of services, user orientation, and efficiency improvement (Arbeids-og Sosialdepartementet [ASD], 2004–2005). The first initiative to reform the welfare administration was taken in 2001 by the Norwegian parliament (Storting) and a final decision was made in 2005, when the reform was adopted by a broad majority of the Storting. At the same time, it was decided to evaluate the reform. The implementation of the reform was initiated in 2006 and lasted until 2010. The reform underwent changes during this time, with two major reorganizations taking place; in 2008, for example, the so-called administrative units were rolled out. The time span of the evaluation was from 2007 to 2014.

The local government reform in Denmark came into force on January 1, 2007. It comprised an amalgamation reform, a task reform, and a financial reform (Indenrigsministeriet, 2005). A total of 271 municipalities were amalgamated into 98, and 14 counties were amalgamated into five regions. The task reform changed the division of labor between state, regions, and municipalities, and the financial reform changed the financing of the tasks. The objectives of the reforms were, as in the Norwegian case, multiple and somewhat ambiguous: to create sustainable local governments, to solve tasks close to citizens, and to secure synergy gains. Major parts of the reform were adopted by

Table 3. Core Elements in the Two Large-Scale Public-Sector Reforms.

| Dimensions | Labor and Welfare Reform (Norway) | Local Government Reform (Denmark) |
|--------------------------------|--|---|
| Reform decision | Decided by Parliament (Stortinget). Short time from adoption to entry into force (2005 to January 1, 2006) | Decided by Parliament (Folketinget). Short time from adoption to entry into force (2004/06 to January 1, 2007) |
| A large administrative reform? | Yes. Implies multiple actors, both structural and other changes: merger of three different units into one unit, partnership organizations (state and municipalities) | Yes. Implies multiple actors, both structural and other changes: amalgamations (from 271 municipalities to 98, from 14 counties to 5 regions), task, and financial reform |
| Multipurpose? | Yes and not very clear: integration of services, user orientation, and efficiency | Yes and not very clear: sustainable local government, solve tasks close to citizens, and synergy gains |
| Evaluations of the reform? | Yes. Decision embedded in reform decision 2005. Evaluation from 2007 to 2014 | Yes, but not decided until 2011 by new government after political disagreement (2007–2011). Evaluation from 2011 to 2012 |

a small majority of the Danish Parliament (Folketinget). The implementation of the reform was initiated in 2005 and spanned over several years. Parts of the reform were changed as it was implemented. For instance, the labor market services were transferred to the municipalities in 2009. An evaluation of the reform was not decided until 2011 after a period with political disagreement on the question and after a new government had come into office.

As stated earlier, the two reforms have on the one hand several significant similarities. They are both large-scale public-sector reforms involving a changed distribution of tasks between the state and local governments. On the other hand, there was a significant difference in the approach to the question of whether or not the reform should be evaluated. Thus, in Norway, the evaluation was embedded in the decision about making the reform, whereas in Denmark the evaluation issue created political disagreement. This difference between the two evaluations will be analyzed further below, with particular focus on whether there are also differences in the three dimensions previously stated in the theoretical framework: (1) the evaluation process (focus and purpose), (2) the evaluators and the organization of the evaluations, and (3) the utilization of the results of the evaluations. The major findings regarding the differences between the two evaluations are summarized in Table 4.

The Evaluation Process: Focus and Purpose

In the evaluation literature, the object of an evaluation can be specified on the basis of whether it focuses on the whole or on individual parts, whether it has a policy or organizational focus, a focus on systems or individual organizations/case studies, or a focus on outcomes or processes. Similarly, the literature introduces a whole range of possible purposes for a given evaluation—for example, checking whether the intentions were realized; documenting effects; identifying implementation problems encountered in order to make adjustments; or enlightening, in terms of broad policy learning (H. F. Hansen, 2005a; Rossi, Freeman, & Lipsey, 2004; Vedung, 1997).

The focus of the NAV evaluation was all parts of the comprehensive reform. Thus, the focus was broad and comprised both the whole and the individual parts, and the policy and organization. Similarly, the purpose of the NAV evaluation was to assess the degree of goal fulfillment (Are more people employed? Has there been an increase in user orientation and efficiency?), to assess the implementation process, to suggest adjustments, and to enlighten through long-term knowledge building. The broad focus and the multiple purposes were determined at an early stage in a “Goal and framework document” (ASD, 2006). It was taken for granted that the reform should be

Table 4. Comparison of Two Evaluations of Large-Scale Public Management Reforms.

| Dimensions | Evaluation of Labor and Welfare Reform (Norway) | Evaluation of Local Government Reform (Denmark) |
|---|--|---|
| The evaluation process (focus and purpose) | Focus: Broad focus on central reform elements established from the beginning Purpose: Goal fulfillment, documentation of implementation process, policy learning | Focus: Initially broad focus, later narrow focus on needs for adjustments in four policy areas, gradual change toward more focus on goal fulfillment Purpose: Formally a “service check,” informally: to legitimize the reform |
| The evaluators and the organization of the evaluation | External: Two consortiums organized into seven modules headed by university and sector researchers | Internal: A committee with stakeholders headed by a high-level civil servant |
| Utilization of evaluation results | Broad knowledge building and democratic debate. Little input to implementation process and adjustment of reform Results: Hardly any improvement in employment or user orientation. Several organizational challenges 87 publications | Recommendations to adjust the reform are under implementation. Reestablished political consensus on the utility of the reform Results: Interorganizational coordination problems. Suggestions to adjust the reform. Probably more sustainable local governments. 13 publications |

evaluated, and both practitioners and researchers took part in the preparation of the document. The Norwegian Research Council organized the evaluation and appointed a steering committee including representatives of the most important players in the evaluation. In accordance with its different purposes, the execution of the evaluation of the NAV reform was organized as seven different subprojects coordinated by a project leader.

One of the important differences between the two evaluations concerns the evaluation process, its focus, and its purpose. The focus and purpose of the LGR evaluation was not made clear from the beginning, but it was developed over time in five phases.

Once the reform process began, the opposition suggested a broad evaluation, but they could not obtain a majority in Parliament (Folketinget, 2009, 2010; Folketingets Lovsekretariat, 2011). In this phase, attempts were made to clarify the object of an evaluation, but the purpose was not specified.

After an election, the opposition took over and an evaluation had become a political commitment: “The government will conduct an evaluation of the local government reform and the current division of responsibilities between municipalities, regions and the state. This process will focus on health, specialized social services, the environment and regional development” (Regeringen, 2011, p. 64). The focus of the evaluation was more precise, but quite broad, focusing on both the whole and the parts and on both policy and organization. The purpose of the evaluation had still not been clarified.

When the terms for the evaluation were published, they stated that the evaluation was to result in an assessment of “. . . the division of responsibilities between municipalities, regions and the state at the moment, and to consider appropriate adjustments to the current situation” (Ministry for Economic Affairs and the Interior [MEI], 2012). Thus, the focus of the evaluation had been significantly reduced. Only if the analysis revealed interface problems, would discussions need to be held about solutions. The focus was no longer on the whole reform, but four policy areas, and it was not on the content of the policy, but on organizational areas (interfaces between parts of the system). The purpose had been narrowed down to a service check (potentially resulting in adjustments) rather than a study of the extent to which the intentions behind the reform had been met or a study of its effects.

During the evaluation process the evaluation committee formulated multisectorial themes on which they wanted to focus (including political governance and sustainability). The themes paved

the way to incorporate a broader focus on the reform. This also opened up for slippage in the purpose in the direction of incorporating an assessment of the degree of goal fulfillment.

When the results of the evaluation were published, the main conclusion (see the Results section) contained an assessment of the degree of goal fulfillment. Thus, the slippage identified above is confirmed by the report.

Thus, the actual structure of the LGR evaluation ended up with having both four evaluations focusing on different areas and an overall evaluation.

The Evaluators and the Organization of Two Evaluations

Evaluation literature distinguishes between internal and external evaluations mainly using organizational boundaries as dividing lines and focusing on whether the producer of the evaluation, the evaluator, is chosen within or outside the organization (Mathison, 2005; Vedung, 1997). Applying this distinction to large-scale public-sector reforms means that if actors responsible for reform implementation are appointed as evaluators, the evaluation becomes internal, whereas if evaluators are appointed from outside the organizational field in which the reform is implemented, the evaluation becomes external. It is of course possible to combine the two types of evaluation.

The NAV evaluation was organized as a special kind of external evaluation. Thus, the Ministry of Labor and Inclusion was responsible for the overall evaluation, while the Norwegian Research Council was responsible for the organization, including the professional and administrative contact with the evaluators in the research environment. The selection of evaluators took place in a process where the research council appointed an expert group to assess applications and propose which to choose to a steering committee consisting of three members appointed by the research council and stakeholders. The appointed evaluators consisted of two consortiums, one of which was given the overall responsibility, while a leader was assigned to each of the seven modules into which the evaluation was organized. The overall budget framework for the period from 2007 to 2014 was approximately NOK 45 million.

The LGR evaluation was internal in the sense that responsibility for it lays with stakeholders also responsible for implementing LGR. They were organized in a relatively restricted committee with four subcommittees, one for each of the four areas. The committee was composed of representatives from key ministries and local government associations. An experienced director general was appointed chairperson. The subcommittees comprised representatives of the same stakeholders as the committee, supplemented by representatives from the relevant sectorial ministry, and with a civil servant from that ministry as chairperson. The relevant sectorial ministry, in collaboration with the MEI and the Ministry of Finance, supported the secretariat function. Overall, the organization of the evaluation reflects the fact that the local government associations were accorded a central position and that the sectorial ministries had a clear platform from which to provide input. In practice, the organizational construction worked by the subcommittees acting as forums for both technical discussions and negotiation and bargaining. Moreover, the main committee held a series of consultations, for instance, with politicians and civil servants from local government and with interest groups. A conservative estimate of the overall cost of this evaluation is more than DKK 10 million, based on the working hours spent by the involved civil servants in ministries, associations, local governments, and interest groups.

Utilization of Evaluation Results in the Two Evaluations

Within the evaluation literature, much attention is devoted to the questions of how evaluations can be used and which type of processes they may influence. Distinctions are made both between different forms of utilization, for instance, an instrumental use, enlightening use, legitimizing use, and tactical use (Fleischer & Christie, 2009; Shulha & Cousins, 1997; Vedung, 1997) and between specific influence processes related to, for instance, cognitive, motivational, and behavioral

processes (Mark & Henry, 2004). The use of the NAV evaluation has primarily been enlightening, for long-term knowledge creation and democratic debate, and only to a limited degree instrumental, to influence the implementation process. Cognitive influence processes seem to be important.

The results of the NAV evaluation were disseminated during the process, for instance, when the evaluators participated in the public debate and in a series of meetings with the stakeholders. The results were documented in roughly 87 publications, most of them with a Norwegian research audience as the target group. Most of the studies focused on the implementation of the reform and only some on the degree of goal fulfillment. Overall, the evaluation resulted in pinpointing a series of more general perspectives, for example, that it was too comprehensive for several municipalities to integrate three services, that it was challenging to establish partnerships between the state and local governments, that the employment of specialists or generalists in the NAV offices was connected with dilemmas, and that the reform (in the short run) hardly had improved job employment, user orientation, or efficiency. The evaluation has only to a limited degree resulted in input to the implementation process and proposals for adjustments. Instead, it contributed to long-term knowledge creation and democratic debate due to the many publications and the dissemination of the results during the process.

The results of the LGR evaluation were documented in 13 publications and a press release from the minister published in March 2013. The main findings were as follows (MEI, 2013, pp. 19–20): “Overall, the committee concludes that the local government reform has created a framework for a more robust public sector that is better able to meet current and future challenges. One of the main goals of the reform was to produce municipalities and regions with greater professional and financial sustainability. Another expectation was that the local government reform would be able to support cost-effective service provision with economies of scale, etc. [. . .] On the basis of its analysis, the committee is of the view that municipalities and regions are in the process of realizing these objectives.” The main conclusion must be said to legitimize the reform. In other words, the trend is in the right direction as defined by the objectives behind the reform. Moreover, the LGR evaluation included many specific assessments of and proposals for each of the four areas, entirely in keeping with the terms of reference. The committee found problems with the interface between municipalities and regions in three areas. In the area of specialized social services, the committee proposed the establishment of a national coordination structure, in which the National Board of Social Services plays a quality assurance role and develops specialist knowledge and services, and that the politicians should choose between three future models: (1) unchanged responsibilities (which the committee clearly does not believe will solve the problems), (2) a municipal model, and (3) regional responsibility for services for specific, limited target groups. When it came to the area of nature and the environment, the committee also proposed different future models, including a future state model, an extended municipal solution, or transferring tasks from municipalities to regions. Concerning rehabilitation (health), the committee suggested a clearer regional competence, guidelines for visitation, and increased transparency.

The LGR evaluation has first of all been used instrumentally to adjust the reform. Behavioral influence processes have been important. Hence, shortly after the evaluation was published, the Danish government presented proposals in June 2013 to adjust the reform according to the recommendations of the evaluation and later that month, broad political agreement was reached (Government, 2013a, 2013b). Moreover, the evaluation has been used to legitimize the reform, as the question of major changes in the reform is no longer on the political agenda. However, in the longer term, the importance of the evaluation depends in part on the ongoing administrative follow-up, and on whether the Liberals, after returning to government in 2015 still want to abolish the regions.

Table 5. Comparison of the NAV and LGR Evaluations Based on Four Theoretical Perspectives.

| Perspective | The NAV Evaluation | The LGR Evaluation |
|-------------------------------------|--|--|
| Rational-instrumental perspective | Contributes to the understanding of the organization and design of the evaluation. Limited contribution to the understanding of the use of the evaluation results, as these were used for enlightenment rather than in an instrumental way | Limited contribution to the understanding of the overall design and the process of the evaluation. Contribution to the understanding of the way in which the secretariat functions of the evaluation were organized. Considerable contribution to the understanding of the political follow-up and use of the evaluation results |
| Rational interest-based perspective | Limited contribution to the understanding of the evaluation. However, some contribution to the understanding of the negotiation processes between the research groups about the division of labor | Considerable contribution to the process ahead of the decision to conduct an evaluation, to the overall organization of the evaluation, and to the process in which conclusions and proposals were framed |
| Institutional-cultural perspective | Contribution to the understanding of the fact that it was natural thing, an SOP, to carry out an evaluation as well as for the taken-for-granted organizational model | Contribution to the understanding of elements in the way in which the evaluation was organized (the corporatist culture) as well as to elements in the design (pragmatism) |
| Chaos perspective | Limited contribution to the understanding of the evaluation. Although some contribution to the understanding of the process in which the four performance areas were agreed on | Considerable contribution to the understanding of the emergent and unpredictable character of the evaluation process where the focus of the evaluation was first narrowed and later enlarged |

Note. NAV = reform of the labor and welfare administration in Norway; LGR = local government reform in Denmark; SOP = standard operating procedure.

Theoretical Interpretations of the NAV and LGR Evaluations

On the basis of the analysis it is discussed below how the differences between the two evaluations can be interpreted according to the four governance perspectives. Table 5 sums up the contribution of the four positions to the interpretation of the two evaluations.

The *rational-instrumental perspective* contributes to various extents to understanding the two evaluations. Whereas it to some extent contributes to understanding the organization as well as the evaluation process in the Norwegian case, it only in a limited way contributes to understanding the Danish case. Hence, instrumental rationality characterized part of the evaluation process in the Norwegian case as the decision to carry through an evaluation was based on an aim to ensure feedback on reform results and contribute to long-term learning and knowledge creation. According to this, it was decided to ask the Norwegian Research Council to take responsibility for arranging a tender in order to be able to select the evaluation project that best suited to fulfill the aims. In 2014, this decision was followed up by a decision to conduct an evaluation of the evaluation, a so-called meta-evaluation, to examine how the public sector can best plan, organize, and carry out public-sector reform evaluations in the future. Contrary to this, neither the process ahead of the decision to conduct an evaluation nor the evaluation process as such, including the overall evaluation design, was characterized in the Danish case by instrumental rationality. It is, however, striking how the opposite pattern is found in relation to how the evaluation results were used in the two countries. In Norway, the evaluation results were used for enlightenment and to initiate debate but not directly for

instrumental decision-making, even though one of the central aims was to secure feedback. In Denmark, by contrast, the evaluation results were used instrumentally in political decision-making, adjusting the reform in an attempt to solve the problems revealed in the evaluation process. In addition, in the Danish case, some elements in the organization of the evaluation can be interpreted as reflecting instrumental rationality. Whereas the organization of the Danish case mostly reflects the interest-based perspective (a theme we elaborate on below), the way of organizing the secretariat function can be interpreted as an instrumental action ensuring strong coordination of the work.

Also, the *interest-based perspective* contributes in varying degrees to understanding the two evaluations. Whereas the perspective contributes substantially to understanding the Danish case, it makes a limited contribution to understanding the Norwegian case. Whereas discussions about whether or not to conduct an evaluation went on in Denmark for years after the political decision on the reform, there were on the contrary no discussions about this in Norway. In the Norwegian context, it was taken for granted that the reform should be evaluated. Likewise, whereas the writing up of reports and conclusions in the Danish case was a negotiation process, it was an analytical process in the Norwegian case. Only in the phase in which the Norwegian research groups were to reach agreement on the division of labor did interest-based negotiation processes seem to have characterized the evaluation process.

The *institutional-cultural perspective* contributes to the understanding of both cases but in relation to different aspects. In the Norwegian case, the fact that conducting an evaluation was taken for granted can be interpreted as a reflection of an institutionalized public-sector evaluation culture (Dahler-Larsen, Nordkvelle, Fosslund, & Nettelund, 2013). In the Danish case, the central stakeholders were invited into the evaluation process in both the committee and the subgroups, which reflects the Danish tradition for public-sector corporatism (H. F. Hansen & Jørgensen, 2009). Likewise, the design of the Danish evaluation can be interpreted as a reflection of the Danish tradition for pragmatism, defined as focusing on what it is practical to accomplish. The fourth perspective, the *chaos perspective*, contributes, as do the first two perspectives to varying extents, to the understanding of the two cases. Whereas it makes a considerable contribution to the understanding of the emergent and unpredictable evaluation process in the Danish case, its contribution to the understanding of the process in the Norwegian case is limited. There is, however, one phase in the Norwegian case, the process in which the four performance areas are agreed on, where ambiguity is high, and coincidences are at play. This phase can be interpreted as reflecting the chaos perspective.

Conclusion

Evaluating large-scale public-sector reforms is a great challenge as different political interests, stakeholders, and target groups are at stake. The purposes are typically ambiguous from the beginning and they can change over time. In this article, we have demonstrated how to reach an understanding of the dynamic and complex nature of the evaluation of such reforms. The analyses of the comparative case study have shown that despite the similarities of the two reforms, their evaluations were organized very differently and given varying importance. First, it was only in Norway that evaluating the reform was taken for granted. In Denmark, there was a political conflict concerning the evaluation for several years. Second, while the evaluation of the NAV reform in Norway was organized as a research-based external evaluation in a process that to some degree was characterized by instrumental rationality, the evaluation of the municipal reform in Denmark was organized as an internal evaluation, in which key stakeholders were given responsibility for the evaluation and thereby a platform to safeguard interests. The evaluation process was more a negotiation process than an analytical process in the Danish context.

The analysis furthermore illustrates how the chosen way to organize the evaluation provides structure and direction to the evaluation process and the utilization of its findings. In the Norwegian case, the external organization produced evaluation findings that were seen by most actors as generally interesting and useful for long-term knowledge building of relevance to the broader society. Findings from the evaluation were disseminated widely in the scientific community and used in the public debate. On the other hand, those responsible for implementing the reform in Norway did not perceive its findings to be of much use in terms of practical guidance concerning how to adjust and improve the ongoing implementation of the reform. In the Danish case, on the contrary, the evaluation did not have much impact on the scientific discourse or the public debate, but the internal organization seems to have been effective in producing proposals for the adjustment of the reform process that were both politically acceptable, possible to implement, and had the support of the actors responsible for implementing them. The political and emerging dimensions of the organization of the Danish evaluation thus provide an important basis for the succeeding instrumental utilization of its proposals.

The two evaluations also differed greatly on the time dimension. While the Norwegian evaluation went on for 7 years, the Danish evaluation was delimited to 1 year. When the decision to evaluate large-scale reforms is taken consensually in the early stage, it is possible to evaluate the entire implementation process with all the challenges of such an undertaking. This happened in the Norwegian case and underlines the interpretational power of the rational-instrumental perspective. However, it also seems to hamper instrumental utilization because large-scale complex reforms often undergo changes as they are implemented over a long time period. On the other hand, when the decision whether or not to conduct the evaluation is a matter of political controversy, as in the Danish case, the rational-instrumental perspective loses some of its interpretational power, due to the unitary features of this perspective. The decision to evaluate is postponed and when it is made several years into the implementation stages of the reform, the patience to wait for evaluation findings may be limited. There is pressure to reach fast conclusions. This happened in the Danish case and stresses the interpretational power of the rational interest-based perspective. But at the same time, it leaves an open playground for negotiations.

The comparative analyses indicate important differences in the roles of the evaluators in the two types of evaluations. The externally organized research-based evaluation tends to enhance a distance between the evaluator and the evaluand; the autonomy of the evaluator and valid knowledge based on scientific standards are the dominant values. The internally organized research-based evaluation with a rather tight time schedule tends to enhance close relations between evaluator and evaluand; the rallying of consensus concerning politically and administratively feasible practical proposals tends to be the dominant value. We do not imply that useful advice was neglected in the Norwegian case or that valid knowledge generation was neglected in the Danish case, but the priorities in the case of trade-offs seem to us to have been pretty clear. We will return to the question about internal or external evaluation and the roles of the evaluators below in the paragraph about implications.

Our analysis has also shown how the institutional-cultural perspective enhances our understanding of the differences between the two countries. In Norway, it was considered obvious and beyond dispute to evaluate the NAV reform. The task was given to the Norwegian Research Council, an organization with a mature evaluation culture. Alternative ways of organizing the evaluation were not considered. In Denmark, on the other hand, it was a matter of political dispute for several years whether or not to evaluate the municipal reform. When the decision to evaluate was finally made, practical considerations of feasibility based on the political agenda and the short time horizon led to an internally organized evaluation. Finally, the chaos perspective contributes to understanding evaluation processes as emergent, rather than linear and predictable. Overall, this perspective is first and foremost relevant to the Danish case.

Applying a theoretical framework including four perspectives has enabled us to better grasp the complexity of evaluating large-scale public-sector reforms, which are always in flux, and characterized by ambitious and ambiguous aims. This was not only illustrated by the fact that the four perspectives to various extents contribute to understanding the different phases of the two evaluations. It is also interesting to observe how instrumental rationality in the planning phase and organization of an evaluation does not necessarily result in instrumental utilization, which was clearly illustrated in the Norwegian case.

Implications for How to Handle Evaluation Challenges

It was also an ambition of ours to answer the difficult question of how to approach evaluations of large-scale public-sector reforms. In that respect, the different perspectives can also be used to illustrate how challenges in evaluations organized as external research or as internal evaluations could be handled. External research-based evaluations, such as the Norwegian NAV evaluation, typically have ambitions about generating generalizable knowledge and spreading this to the academic community (as well as the evaluation sponsor), as such evaluations develop in an academic culture that poses basic questions and critically assesses methodological issues (Halvorsen, 2013, p. 244). In this context, it is a challenge to obtain evaluation results that can be considered usable in the praxis of implementing and further developing the reform in focus. However, the interest-based perspective, the institutional-cultural perspective, and the chaos perspective hopefully can give some ideas about how to handle these challenges. According to the interest-based perspective, external evaluators, in addition to considering themselves academics, could play the role of leaders of negotiations, inviting stakeholders into a dialogue, and listening to stakeholders' demands for knowledge. According to the institutional-cultural perspective, actors responsible for reform implementation and development must be expected to have limited confidence in the ability of academic evaluators to deliver usable results. Thus, academic evaluators have to work consciously with building trust. Finally, according to the chaos perspective, academic evaluators could try to create platforms to facilitate processes and to discuss and handle reform ambiguity.

In contrast, internal evaluations can be expected to be at risk of producing results that are not considered trustworthy and independent, but rather to be contributions supporting powerful stakeholders. In this context, internal evaluators could use the rational-instrumental perspective as a model for putting methodological issues on the agenda for critical discussion in a strategy to enhance the credibility of results. Moreover, the interest-based perspective can generate reflections on how to reach a compromise that can be expected to be lasting, and according to the institutional-cultural perspective, reflections can be made on how to create a trustworthy evaluation, also for others than the involved stakeholders. Finally, according to the chaos perspective, internal evaluators could also try to create platforms to facilitate processes and to discuss and handle reform ambiguity.

Questions for Further Research

Our comparative analysis above has uncovered important characteristics related to the evaluation of large-scale public-sector reforms and indicated how evaluators may cope with the challenges posed by such evaluations. Like most research projects, the analysis also raises a number of questions which we can only scratch the surface of and which we must leave as open questions for future research.

One question posed by our analysis is why large-scale public-sector reforms are a rare evaluand and why they, despite that, were evaluated in our two cases? In the introduction and the literature review, we established that such reforms are rarely evaluated, especially in terms of external evaluations, and used it as rationale for the importance of our study. It is however a puzzle, akin to a "paradox of public management reform" (Pollitt & Bouckaert, 2011b, p. 158), that large-scale

public-sector reforms often emphasizing performance management and evaluation as essential governance instruments are rarely evaluated themselves. A recent study, finding negative or very little impact of the New Public Management reforms in the United Kingdom in recent decades (Hood & Dixon, 2015), may indicate a partial answer. In order to mobilize support, such reforms are often “oversold” and their results are thus likely to be disappointing by reformers own self-imposed standards. Since such reforms are irreversible, their evaluation is hardly appealing to responsible reformers. Nevertheless, in our cases, the two reforms were evaluated and in the Norwegian case this evaluation was considered self-evident. This poses questions concerning which contingencies may enhance or inhibit the evaluation of large-scale public-sector reforms, which is however beyond the scope of the present analysis.

A second question, by and large excluded from our analysis, is related to the technical quality of the evaluations. An obvious (probably too obvious) thesis might be that external research-based evaluations are more likely to enhance a high level of technical quality in the evaluation than internal evaluations. Technical quality is a tricky issue though and depends on purpose and as well as a number of other contingencies.

Space only allows us to raise these final intriguing questions, while possible answers are left to future research projects.

Acknowledgments

The authors would like to thank Editor Christina Christie and two anonymous reviewers for important suggestions for improvement of earlier drafts. We would also like to thank professor Peter Dahler-Larsen for helpful comments on an earlier Danish version of the paper.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article was supported by The Norwegian Research Council.

Note

1. For a few studies on reforms in for example Finland, New Zealand, and Sweden, see H. F. Hansen (2005), Pollitt and Finnland (1997), Thoenig (2000), and Wollmann (2003a).

References

- Allison, G. T. (1971). *Essence of decision. Explaining the Cuban missile crisis*. Boston, MA: Little, Brown and Company.
- Andreassen, T. A., & Aars, J. (2015). *Den store reformen. Da NAV ble til*. Oslo, Norway: Universitetsforlaget. Arbeids-og Sosialdepartementet. (2004–2005). *St. prp. nr.46: Ny arbeids- og velferdsforvaltning*. Oslo, Norway: Author.
- Arbeids-og Sosialdepartementet. (2006). *Mål og rammedokument for evaluering av NAV-reformen*. Oslo, Norway: Author.
- Breidahl, K. N., Furubo, J.-E., Halvorsen, K., & Hansen, M. B. (2014). *Opmerksomhetspunkter ved evaluering af politiske reformer* (Rapport fra metaevalueringen af NAV-reformen i Norge). Oslo, Norway: The Norwegian Research Council.

- Breidahl, K. N., Gjelstrup, G., Hansen, M. B., & Hansen, H. F. (2015). Evaluering af politisk-administrative reformer: En komparativ analyse af evalueringen af de største reformer i Danmark og Norge. *Økonomi og Politik*, 88, 34–45.
- Broadbent, J., & Laughlin, R. (1997). Evaluating the 'new public management' reforms in the UK: A constitutional possibility? *Public Administration*, 75, 487–507.
- Brochmann, G., & Hagelund, A. (2012). *Immigration policy and the Scandinavian welfare state 1945–2010*. Basingstoke, England: Palgrave Macmillan.
- Dahler-Larsen, P., Nordkvelle, I., Fosslund, T., & Nettelund, G. (2013). Kvalitetens beskaffenhed og den svage tænkning. *Kvalitet i fleksibel højere uddanning—nordiske perspektiver*, 29, 29–40.
- Fimreite, A. L., & Lægred, P. (2009). Reorganizing the welfare state administration. *Public Management Review*, 11, 281–297.
- Fleischer, D. N., & Christie, C. A. (2009). Evaluation use results from a survey of US American evaluation association members. *American Journal of Evaluation*, 30, 158–175.
- Folketinget. (2009). *Forslag til vedtagelse V8 Om regioner og kommuner, 11.11.09*. København. København, Denmark: Folketinget, the Danish Parliament.
- Folketinget. (2010). *Forslag til vedtagelse V69 Om den kommunale økonomi, 07.05.10*. København. København, Denmark: Folketinget, the Danish Parliament.
- Folketingets Lovsekretariat. (2011). *Svar til medlem af Folketinget Julie Skovsby på spørgsmål nr. S 2025 til indenrigs- og sundhedsministeren*. København, Denmark: Folketinget, the Danish Parliament.
- Gjelstrup, G., & Hansen, H. F. (2014). Evalueringsstyring i en krisetid—kommunalreformen. In C. H. Grøn, H. F. Hansen, & M. B. Kristiansen (Eds.), *Offentlig styring. Forandringer i krisetider* (pp. 385–416). København, Denmark: Hans Reitzels Forlag.
- Government. (2013a). *Aftale mellem regeringen, Venstre, Dansk Folkeparti, Enhedslisten, Liberal Alliance og Konservative, Rammer for justering af kommunalreformen, 26. januar*. Copenhagen, Denmark: The Danish Government.
- Government. (2013b). *Bedre kvalitet og samarbejde—Opfølgning på evaluering af kommunalreformen, juni*. Copenhagen, Denmark: The Danish Government.
- Grøn, C. H., Hansen, H. F., & Kristiansen, M. B. (2014). Kultur-, symbol- og kaosperspektiver på styring. In C. H. Grøn, H. F. Hansen, & M. B. Kristiansen (Eds.), *Offentlig styring. Forandringer i krisetider* (pp. 131–160). København, Denmark: Hans Reitzels Forlag.
- Halvorsen, A. (2013). Forskningsbaserede evalueringer og andre evalueringer. In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering. Traditioner, praksis, mangfold* (pp. 233–246). Bergen, Norway: Fagbokforlaget.
- Hansen, H. F. (2005a). Choosing evaluation models. *Evaluation*, 11, 447–462.
- Hansen, H. F. (2005b). Evaluation in and of public-sector reform: The case of Denmark in a Nordic perspective. *Scandinavian Political Studies*, 28, 323–347.
- Hansen, H. F. (2011). NPM in Scandinavia. In T. Christensen & P. Lægred (Eds.), *New public management* (pp. 113–130). Farnham, England: Ashgate.
- Hansen, H. F., & Jørgensen, T. B. (2009). Den danske forvaltningsmodel og globaliseringens udfordringer. In M. Marcussen & K. Ronit (Eds.), *Globaliseringens udfordringer. Politiske og administrative modeller under pres* (pp. 36–64). København, Denmark: Hans Reitzels Forlag.
- Hansen, M. B., Breidahl, K. N., Halvorsen, A., & Furubo, J.-E. (2015). Otte opmærksomhedspunkter ved evaluering af politiske reformer. *Økonomi og Politik*, 88, 24–33.
- Hansen, M. B., Breidahl, K. N., Furubo, J.-E., & Halvorsen, A. (In Press). Eight attention points when evaluating large-scale public sector reforms. In B. Greve (Ed.), *Handbook of social policy evaluation* (pp. 460–472; Chapter 23). London, England: Edward Elgar Publishing.
- Hansen, M. B., Lægred, P., Pierre, J., & Salminen, A. (2012). Comparing agencification in Nordic countries. In S. van Thiel, K. Verhoest, G. Bouckaert, & P. Lægred (Eds.), *Government agencies: Practices and lessons from 30 countries* (pp. 259–265). Hampshire, England: Palgrave MacMillan.

- Hansson, F. (1998). Evaluation research and sociology in Denmark: A tale of two cultures. *New Directions for Evaluation 1998* (77 Special Issue: Scandinavian Perspectives on the Evaluator's Role in Informing Social Policy), 55–70.
- Haug, P. (1998). Linking evaluation and reform strategies. *New Directions for Evaluation 1998* (77 Special Issue: Scandinavian Perspectives on the Evaluator's Role in Informing Social Policy), 5–20.
- 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, England: Oxford University Press.
- Indenrigsministeriet. (2005). *Kommunalreformen—kort fortalt*. København, Denmark: Indenrigs- og Sundhedsministeriet.
- Jacob, S., Speer, S., & Furubo, J.-E. (2015). The institutionalization of evaluation matters: Updating the international atlas of evaluation 10 years later. *Evaluation*, 21, 6–31.
- Kuhlmann, S., Bogumil, J., & Grohs, S. (2008). Evaluating administrative modernization in German local governments: Success or failure of the “new steering model”? *Public Administration Review*, 68, 851–863.
- Kuhlmann, S., & Wollmann, H. (2011). The evaluation of institutional reforms at sub-national government levels: A still neglected research agenda. *Local Government Studies*, 37, 479–494.
- Kuhlmann, S., & Wollmann, H. (2014). *Introduction to comparative public administration. Administrative systems and reforms in Europe*. Cheltenham, England: Edward Elgar.
- March, J. G. (1999). Understanding how decisions happen in organizations. In J. G. March (Ed.), *The pursuit of organizational intelligence* (pp. 13–38). New York, NY: Blackwell Publishers.
- March, J. G., & Olsen, J. P. (1986). Garbage can models of decision making in organizations. In J. G. March and R. Weissinger-Baylon (Eds.), *Ambiguity and command: Organizational perspectives on military decision making* (pp. 11–35). Cambridge, MA: Ballinger.
- March, J. G., & Olsen, J. P. (1989). *Rediscovering institutions: The organizational basis of politics*. New York, NY: The Free Press.
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10, 35–57.
- Mathison, S. (Ed.). (2005). *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage.
- Ministry for Economic Affairs and the Interior. (2012). *The local government reform get a service check, press release, 9. February*. Copenhagen, Denmark: Author.
- Ministry for Economic Affairs and the Interior. (2013). *Evaluation of the local government reform*. Copenhagen, Denmark: Author.
- Pedersen, L., & Rieper, O. (2008). Is realist evaluation a realistic approach for complex reforms? *Evaluation*, 14, 271–293.
- Pollitt, C. (1995). Justification by works or by faith?: Evaluating the new public management. *Evaluation*, 1, 133–154.
- Pollitt, C., & Bouckaert, G. (2011a). *Public management reform. Comparative analysis of new public management, governance and the Neo-Weberian State* (3rd ed.). Oxford, England: Oxford University Press.
- Pollitt, C., & Bouckaert, G. (2011b). Results: Through a glass darkly. In C. Pollitt & G. Bouckaert (Eds.), *Public management reform. A comparative analysis: New public management, governance and the Neo-Weberian State* (3rd ed., pp. 126–160). Oxford, England: Oxford University Press.
- Pollitt, C., & Finnland, V. (1997). *Trajectories and options: An international perspective on the implementation of Finnish public management reforms*. Helsinki, Finland: Ministry of Finance.
- Regeringen. (2011). *Et Danmark der står sammen. Regeringsgrundlag oktober 2011*. København, Denmark: Author.
- Roness, P. (1997). *Organisasjonsendringar. Teoriar og strategiar for studiar av endringsprosessar*. Bergen, Norway: Fagbokforlaget.
- Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2004). *Evaluation. A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.

- Schein, E. H. (1985). *Organizational culture and leadership*. San Francisco, CA: Jossey-Bass.
- Schwandt, T. A. (1998). Editor's Notes. *New Directions for Evaluation 1998* (77 Special Issue: Scandinavian Perspectives on the Evaluator's Role in Informing Social Policy), 1–4.
- Scott, W. R., & Davis, G. F. (2007). *Organizations and organizing. Rational, natural and open system perspectives* (Pearson International Edition ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, 18, 195–208.
- Stacey, R. (2011). *Strategic management and organizational dynamics. The challenge of complexity*. Englewood Cliffs, NJ: Prentice-Hall.
- Thoenig, J.-C. (2000). Evaluation as usable knowledge for public management reforms. *Evaluation*, 6, 217–229.
- Van Eyk, H., Baum, F., & Blandford, J. (2001). Evaluating healthcare reform: The challenge of evaluating changing policy environments. *Evaluation*, 7, 487–503.
- Vedung, E. (1997). *Public policy and program evaluation*. New Brunswick, NJ: Transaction Publishers.
- Weick, K. E. (2001). *Making sense of organizations*. Malden, MA: Blackwell Publishers.
- Wollmann, H. (2003a). Evaluation in public sector reform: Trends, potentials and limits in international perspective. In H. Wollmann (Ed.), *Evaluation in public policy reform* (pp. 231–258). Cheltenham/Northampton, England: Edward Elgar.
- Wollmann, H. (Ed.). (2003b). *Evaluation in public-sector reform: Concepts and practice in international perspective*. Cheltenham, England: Edward Elgar Publishing.