The organisation of evaluations: the influence of the ministry of finance on evaluation systems

Andersen, N.A.; Pattyn, V.E.

Evidence & Policy • vol XX • no XX • 1–23 • © Authors 2025
Print ISSN 1744-2648 • Online ISSN 1744-2656 • https://doi.org/10.1332/17442648Y2025D000000045
Accepted for publication 21 January 2025 • First published online 20 February 2025

research article

The organisation of evaluations: the influence of the ministry of finance on evaluation systems

Niklas A. Andersen, nia@dps.aau.dk
Aalborg University, Denmark

Valérie Pattyn, v.e.pattyn@fgga.leidenuniv.nl
Leiden University, The Netherlands

Background: Despite increasing scholarly interest in the organisation of evaluations within different countries' political-administrative landscapes, not much attention has hitherto been paid to the consequences of a specific institutional set-up for the function of evaluations within government.

Aims and objectives: This article investigates how the organisational anchorage of policy evaluations within central administration shapes the function those evaluations primarily serve.

Methods: We focus on the role of ministries of finance for coordinating countries' evaluation systems, and study its influence in Denmark and the Netherlands through a combination of document analysis and interviews with centrally placed civil servants.

Findings: Our analysis shows how the ministries of finance come to influence the evaluation activities of the whole central administration by constituting a specific economic outlook on evaluation, which (1) narrows down the applied evaluation methods and criteria; (2) inserts the ministry of finance as primary evaluation user; and hereby (3) furthers accountability rather than learning as the main function of evaluations within central administration. In both countries, the result is that the ministry of finance's main role in the evaluation systems favours somewhat defensive qualities, where evaluations are primarily used for control and piecemeal changes in policies, rather than fundamental revisions or reflections on the appropriateness of specific policies.

Discussion and conclusions: Our findings indicate that the influence of evaluation systems is not only dependent on the degree of institutional anchorage of evaluation activities, but also very much a matter of *whom* the evaluation systems is centred around.

Keywords policy evaluations • evaluation systems • constitutive effects • ministry of finance

To cite this article: Andersen, N.A. and Pattyn, V. (2025) The organisation of evaluations: the influence of the ministry of finance on evaluation systems, *Evidence & Policy*, Early View, DOI: 10.1332/17442648Y2025D000000045

Introduction

In recent decades, policy evaluations have evolved into a critical element of good governance, given their potential to provide relevant knowledge about the content, implementation or impact of ongoing or completed policies (OECD, 2020). Still, governments can anchor policy evaluations in different ways in their government apparatus, as evidenced by several mapping exercises of prevalent institutional designs (Jacob et al, 2015; OECD, 2020; Stockmann et al, 2020). Despite increasing scholarly interest in the organisation of evaluations in a country's political-administrative landscape, less attention has hitherto been paid to the consequences of a specific institutional set-up for the function of evaluations within government. By function we refer broadly to the many different forms of evaluation use highlighted by the burgeoning evaluation utilisation literature (King and Alkin, 2018). This literature has found a plethora of different types of both intended and unintended usages, such as instrumental use of evaluative findings to design policy programmes (Weiss, 1998) or strategic use of evaluations to bolster political positions and arguments (Boswell, 2009). In the current article we focus on the two primary functions of accountability (or control) and learning (or improvement), which are often highlighted as the main dual purpose of evaluations (Alkin and King, 2017; Reinertsen et al, 2022). Most types of intended use can be categorised into either of these. Both accountability and learning are broad and vague concepts, but here we use them in relation to the prevalent distinction within the evaluation literature between summative and formative use of evaluations (Vedung, 2017). When evaluations serve the function of accountability, they are thus used to pass judgement on that which is evaluated (be it an organisation, an individual, a policy, and so on) and thus ultimately control this through the rewards and punishments stemming from this judgement. When evaluations serve the primary function of learning, they are rather used to further develop that which is evaluated and thus ultimately try to improve this by changing practice according to the knowledge attained. The function at stake is important. In the case of control, evaluations can be viewed as 'information' about results, providing evidence that the government is (or is not) working effectively. In the case of learning, evaluations are more about 'knowledge' that is explanatory and theoretical, which can then be used formatively to drive policy and societal change. Depending on the evaluation function, the type of evidence is therefore qualitatively different in nature (Sanderson, 2002). Both functions are legitimate within government, with neither being superior to the other; however, existing research indicates that the two functions can be in tension with each other (Reinertsen et al, 2022). The factors furthering one of the functions may impede the other function.

The current article investigates this potential tension by elucidating the influence of organisational factors for the function of evaluations within central administration. The article seeks to answer the question of how the organisational anchorage of evaluations within the ministry of finance influences the functions evaluations primarily serve within policy making. We focus on the role of ministries of finance as this is the most common organisational anchorage for the coordination of evaluation activities within central administration in Organisation for Economic Co-operation and Development countries (OECD, 2020). Empirically, we focus on Denmark and the Netherlands, which both have a high degree of evaluation maturity in central government (Jacob et al, 2015) and where the ministry of finance has played a central role in the coordination of the evaluation function for many years.

With the emphasis on the organisational dimension of evaluations and the specific role of the ministry of finance, we contribute to scholarship on evaluation use by bridging two hitherto disparate literatures. The first is the small but growing literature on evaluation systems (Rist and Stame, 2006; Leeuw and Furubo, 2008; Raimondo, 2018; Kupiec et al, 2023b), which advocates a shift away from focusing on the influence of individual evaluation studies towards a more holistic emphasis on the very context in which evaluation studies are being produced. Studies on evaluation systems have generally focused on elucidating the processes of institutionalisation through which evaluation systems have come about, as well as analysing how such systems affect the use of evaluations (Højlund, 2014; Andersen, 2020). However, the literature has hitherto paid less attention to unpacking the relative importance of specific dimensions of the institutionalisation and systematisation of evaluation activities. That is, whether and how the specific institutional set-ups and the role and power of the central actors within an evaluation system further certain evaluation functions rather than others.

To fill this lacuna, we draw on a second strand of literature concerning the power and influence of economists in central government (Hirschman and Berman, 2014). This literature offers a perspective that – compared to the literature on evaluation systems – is more attuned to the importance of specific actors and interests for how evaluation systems are structured and utilised. Existing studies have aptly shown how economists in general have gained an increasingly central role in policy making over the last 30–40 years (Christensen, 2018; Christensen and Mandelkern, 2022). This development is closely related to a subsequent change in the role of the ministries of finance, which – across many different countries – have strengthened their position of power vis-à-vis other ministries since the 1980s (Roberts, 2010; Garritzmann and Siderius, 2024). This strengthening of the power of the ministry of finance, and its economists, have come about both through a general centralisation of government policy making (Raudla et al, 2015; Kristiansen, 2018) and through the introduction of policy instruments such as fiscal rules (Roberts, 2010) or new economic models (Fuglsang, 2023). Studies have shown how this has infused government policy making with certain logics (Maron, 2021; Garritzmann and Siderius, 2024) or styles of reasoning (Heimberger et al, 2020; Berman, 2022) focused on efficiency and frugality. Our article draws on such findings to elucidate the way in which the economic thinking of the ministry of finance is inscribed within the evaluation system of the central administration and thus affects the way evaluations of current and prior policies are structured and used.

The article proceeds as follows: in the next section, we present our theoretical framework concerning the 'constitutive effects' of evaluation systems. We then proceed with a presentation of our methodological approach, after which we zoom into the findings. We conclude with a discussion section in which we discuss the trends and differences across both country cases and the broader implications of our findings.

The constitutive effects of institutional arrangements: theoretical approach

Leeuw and Furubo (2008) argue that in many modern organisations, evaluation activities are increasingly becoming 'more permanent arrangements' (p 159), where single evaluations are embedded within broader evaluation systems. According to the authors, such evaluation systems create coherence between the different evaluation activities of a given organisation or within an organisational field through four central dimensions: (1) ensuring that evaluations are produced on a permanent basis and with a substantial volume; (2) endowing one or more central organisational units within a given organisational field with the responsibility for the commissioning and/or production of evaluations; (3) ensuring that evaluation activities are guided by a shared epistemological perspective; and (4) ensuring that procedures are in place to enhance the instrumental use of evaluative findings.

Leeuw and Furubo thus argue that in situations where each of these dimensions are met, it makes sense to study evaluation activities as part of a coherent evaluation system. It is exactly through this overarching focus on the coherence of evaluation activities – and how organisational and institutional features can further such coherence – that the concept of evaluation system differs from related concepts such as evaluation culture (Furubo et al, 2002; Jacob et al, 2015) or evaluation capacity (Preskill and Boyle, 2008; Bourgeois and Cousins, 2013). The point being, that it is possible for an organisation to show a high maturity in evaluation culture or have a significant capacity for making evaluations, without having an evaluation system – as defined by these criteria. This is, for example, the case, if the different evaluation activities of the organisation are driven by a dispersed set of actors or draw on different epistemological perspectives.

It is clear from this definition of evaluation systems that the coherence and function of any evaluation system is largely determined by its organisational features – that is, the structures, procedures and processes put in place to support the permanence, methodological rigour and instrumental use of evaluations. Even so, the empirical literature on the systematisation and institutionalisation of evaluations, which has expanded significantly since Leeuw and Furubo's seminal article, shows surprisingly little systematic reflection on the association between the specific institutionalisation modalities or ways of organising evaluations in a given political-administrative setting and evaluation use in that same setting. The lack of such nuanced studies is symptomatic for the general absence of discussion about different forms of institutionalisation with their advantages and disadvantages (Jacob et al, 2015). A range of studies indeed focus on organisational determinants influencing evaluation use (see Kupiec et al, 2023a). While valuable, most of these focus on the meso-level of analysis, which deals with the organisation of evaluation systems within certain ministries or agencies. This leaves aside the more macro-organisational set-up at country level, in which ministries and agencies operate, and which thus influences the meso-organisational level. As already indicated in 2015 (Jacob et al, 2015), little systematic comparative research across countries exists. However, we know from the few systematic comparative cross-country studies (Jacob and Varone, 2004; Filgueiras and Queiroz, 2021; Kupiec et al, 2023b) that different evaluation styles (De Francesco and Pattyn, 2021) and levels of evaluation maturity (Furubo et al, 2002) exist across countries, and that different institutional design scenarios have been adopted (for example, Jacob et al, 2015; OECD, 2020; Stockmann et al, 2020). While there are several parameters defining the institutional set-up of policy evaluations at country level (Furubo et al, 2002; Jacob et al, 2015; Stockmann et al, 2020), the key feature is arguably the organisation providing strategic direction for evaluation within the executive (that is, criterion 2). Recent reflections on evaluation institutionalisation and its importance in view of evaluation capacity building also highlight the prominent role of the executive branch (Lamarque, 2023). This makes it one of the most logical empirical entry points to understand how institutional arrangements shape the use of evaluative evidence within countries.

To study the consequences of institutional arrangements, we rely on the concept of *constitutive effects* (Dahler-Larsen, 2014). The concept of constitutive effects or consequences is indebted to notions of power that centres on the 'productive' aspects of power and knowledge (Foucualt, 1990; Haugaard, 2010), which for long has played a prominent role within sociologically inspired studies on tools and practices of measuring (Espeland and Sauder, 2007), quantifying (Berman and Hirschman, 2018; Mennicken and Espeland, 2019) and accounting (Hopwood and Miller, 1994; Power, 1997). The concept elucidates how tools and practices of knowledge production not only *represent or describe* social reality, but also *constitute or 'make up'* social reality. Through the works of Peter Dahler-Larsen (2012; 2014) the concept has also gained traction within the fields of evaluation and performance measuring in the last decade. Here the concept has been used to challenge the dominant focus on intended and unintended consequence of evaluations and performance measurement, which is closely tied to the preoccupation of both research fields with enhancing instrumental use of evaluative findings and performance information (Alkin and King, 2017). The concept of constitutive effects discards the primacy of any preconceived and fixed intentions of use by the ones who commissioned or produced the evaluation or measurement. Instead, it directs our attention towards the myriad of ways that the act of evaluating and measuring shapes (that is, constitutes) that which is being evaluated and measured – regardless of any formally stated intentions (Dahler-Larsen, 2014).

As such, the concept functions as a sensitising concept, which points to two distinct avenues where evaluations and measurements – through the way they categorise, commensurate, temporalise and fixate the object being evaluated and measured (Dahler-Larsen, 2024) – can come to constitute social reality: the shaping of language and the shaping of social interactions among the central actors of a given organisational field (Dahler-Larsen, 2012). In the instance of language, the categorisation and fixation of meaning within evaluations (for example, how 'quality', 'goals' or 'good performance' is operationalised) can come to colonise the language of the wider organisational field regarding how to view and describe the tasks and goals of organisations within this field. Social interactions are shaped through processes of interpellation where the actors within the broader organisational field come to see themselves and each other through the values, norms and criteria of the evaluation activities. This can, for example, happen by evaluations and measurements inserting standards of high- and low-performing organisations or individuals, which these organisations and individuals then come to see themselves and each other in accordance to.

Dahler-Larsen's conceptualisation of constitutive effects has been used to study the consequences of evaluations and measurements on public sector governance (Hanberger, 2016; Hanberger and Lindgren, 2019; Kok et al, 2019), users of public services (Carlbaum, 2016), welfare professionals (Lindgren, 2015; Hult and Edström, 2016) and processes of policy making (Andersen, 2020). In all instances, the focus of the studies is on the way the production and use of evaluations and measurements constitute a certain way of governing, making policies, viewing citizens and professionals, and so on. In the current article we instead propose to shift focus to the question of how the way evaluation activities are *organised* constitute certain evaluation functions. We do so by studying how the organisational anchorage of the Dutch and Danish evaluation systems – ascribed to the ministry of finance through both formal and informal means – reshapes the language concerning how evaluations are produced and the social relations among the actors ascribed as the primary producers and users of such evaluations. Ultimately we are interested in elucidating whether and how this shaping of language and social relations can come to constitute a dominant understanding of the main function of evaluations – for example, whether evaluations should primarily serve purposes of accountability and control or learning and improvement.

Case countries: Denmark and the Netherlands

Our study focuses on Denmark and the Netherlands. Both countries share three important features with relevance to our research question. First, they exhibit similar characteristics in terms of their politico-administrative tradition. Both are part of the family of parliamentary democracies, have a corporatist-consensus policy-making style, and are considered as 'early modernisers' with high scores on New Public Management emphasis in comparative studies (Hood, 1995; Verhoest, 2016; Lægreid, 2017; Pollitt and Bouckaert, 2017). Additionally, both countries demonstrate relatively low degrees of politicisation (Peters and Pierre, 2004). They also rank among the highest worldwide in 'government effectiveness', as measured by the Worldwide Governance Indicators (Denmark having percentile rank 98.58, the Netherlands having rank 95.28 in 2022) (Lægreid, 2017: 83; Kaufmann and Kraay, 2023). Government effectiveness refers to the perception of the quality of public services, the quality of the civil service, the degree of independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies (Kaufmann and Kraay, 2023).

Second, both countries are consistently at the top of indices of evaluation institutionalisation in Europe (Jacob et al, 2015; Stockmann et al, 2020). They have been showing a permanence and volume in policy evaluations, and have taken substantial efforts to enhance the epistemological coherence and instrumental use of policy evaluations as well as to organisationally anchor the evaluation system within the ministry of finance. They can thus both be classified as highly institutionalised evaluation systems (Leeuw and Furubo, 2008).

Third, and important, in both countries the drive to make policy evaluations a more permanent – and thus systematic – part of ministerial practices stemmed from economic downturns. The Dutch evaluation system received its initial impulses as early as the 1970s, following the stagflation crisis, with efforts to achieve a more goal-oriented budget and a range of spending reviews during the 1980s (Oudenampsen and Mellink, 2022). Similarly, the Danish evaluation system began to take shape after the record levels of unemployment and huge deficits on the balance of payment that plagued the country throughout the 1980s and into the early 1990s. For the ministry of finance, this economic crisis highlighted the need to assume the role of central coordinator and final arbiter of all government policies (Ejersbo et al, 2014). As part of this role, the ministry also took responsibility for developing a common framework to evaluate the costs of policy proposals and to validate and double-check the calculations, evaluations, and projections made by the line ministries responsible for a given policy. This is comparable to the Netherlands, where the ministry of finance has long viewed its main task as ensuring budgetary discipline, with public choice theory serving as an important intellectual foundation (Oudenampsen and Mellink, 2022). With Vedung (2010), we can state that both countries are part of the 'neo-liberal wave', or 'new public management wave' in which evaluations have spread across the European continent (Vedung, 2010; De Francesco and Pattyn, 2022). This wave, which has left its sediments until today (Capano et al, 2022), came with a strong emphasis on 'value for money', and a central role of evaluations in view of results-based management, and accounting for the economy, effectiveness and cost-efficiency of public spending.

In this regard, it is also relevant to refer to the fiscal governance tradition to which both countries belong. They are clear cases of the so-called 'contract approach', where the budget process hinges on 'pre-established, numerical budgetary targets negotiated among key policymakers' (Hallerberg et al, 2006: 2), a practice typical for coalition governments. Under this approach, the finance minister has strong monitoring capacities regarding budget execution and the authority to correct deviations from it.

These three features combined – their shared administrative tradition, the strong institutionalisation of the evaluation system, and the decade-long role of the ministry of finance to further the evaluation system – makes the cases very comparable. However, both countries differ in one important respect, that is the formality of their evaluation system. Whereas the Dutch system has enacted a formal regulatory framework for the production and use of evaluations, the Danish system is based much more on norms, practices and procedures not stipulated by law. Moreover, in Denmark, these are largely focused on *ex ante* evaluations only.

In the Netherlands, the Government Accounts Act (2016) and the associated Regulation for Periodic Evaluation (Ministerie van Financiën, 2022) constitute the regulatory corpus of the evaluation system. In terms of *ex ante* evaluation, Article 3.1 of the Act prescribes that proposals, intentions and commitments must include an explanation detailing (1) the goals, effectiveness, and efficacy that are aimed for; (2) the policy instruments that are used; (3) the financial consequences for central government and, where possible, the financial consequences for social sectors. Proposals that imply 'substantial policy changes' with a budgetary impact of over 20 million euros also need to include an evaluation paragraph, detailing how the policy will be evaluated during and/or after its implementation.

The Act is also important for *in itinere* and *ex post* evaluation. The Act (Article 4.1) mandates that responsible ministers periodically examine the '(conditions for the) effectiveness and efficiency' of their policies and related expenditures. The Regulation for Periodic Evaluation elaborates on this, and specifies how ministers must account for the money they spend on the objectives outlined in departmental budgets (Ministerie van Financiën, 2022). A key instrument in this regard are the periodic reviews. These require ministries to compile the insights obtained from individual evaluations on each (policy) theme of their Strategic Evaluation Agenda into a periodic report at least every four to seven years. This includes, among other requirements, describing opportunities for increasing the policy's effectiveness and efficiency with the same level of financial resources. Additionally, at least one efficient option must be outlined to achieve a 20 per cent budgetary savings for the (policy) theme.

While individual ministers are primary responsible for evaluating their policies, the minister of finance has a strong regulatory and supervisory role (Government

Accounts Act, Article 4.1). It will be this ministry monitoring compliance with the Accounts Act, implying that it may reject a draft budget or amendment to the extent that it is not consistent with the general financial policy or the efficient management of the financial resources of the Kingdom. Similarly, if proposals to spend money 'do not bear a reasonable relationship or are not in reasonable proportion to the objectives of the policy underlying that budget', the finance minister can object (van der Knaap et al, 2020). In sum, in the Netherlands, the evaluation system is very formalised, with the ministry of finance having a central coordination role in this regard.

In contrast, Denmark has taken a less formalised route to enhance instrumental use of evaluations within central administration, as there are *no* formal requirements stipulated by law to commission, produce or use policy evaluations in the policy-making process – neither before nor after a policy is adopted. The Danish evaluation system is thus much more dependent on certain norms and procedures that have been formed over time – rather than being inscribed into formal regulations of the policy-making process. Such norms and procedures have especially developed after the introduction of a new budget law in 2012. With this budget law, it was decided that when ministries develop and propose new policies, the projected cost of these policy proposals had to be covered either by the same proposal or by other policies that were already passed and implemented. Around the same time a new norm began to emerge, where the ministry of finance in select instances started to include not only direct costs, but also more dynamic – or behavioural – effects on the labour supply in the calculation of the expected costs of a policy proposal (Andersen, 2020; Kelstrup, 2024). Until the beginning of the 2010s the norm was that projections on such dynamic effects on labour supply were routinely done by the ministry of finance, but the calculated gains of such an increased labour supply would only be used when and if they were actually materialised. This has changed markedly in the last decade, where expected future gains can be used as soon as a policy is adopted and not just when (and if) these gains are materialised. This has then of course greatly increased the potential use of *ex ante* evaluations that can document expected dynamic effects on the labour supply of a proposed policy. At the same time, this has also further bolstered the epistemological coherence of the evaluations being done by different ministries, as these are increasingly focused on calculating the dynamic effects of policies on the labour supply according to the principles laid down in several guidelines by the ministry of finance (for example, Finansministeriet, 2018; 2021). This gives the ministry of finance an important anchoring role in setting the methodological standards of the evaluation system, which is further bolstered by the procedure, that all policies with expected effects on the labour supply have to first pass through the economic committee, where the ministry of finance is permanent chair. Thus, affording the ministry of finance with an important gatekeeper function concerning policy making.

Taken together, the many similarities in the two contexts, combined with the central difference regarding the formalisation of the organisational anchoring of the evaluation systems, provides a unique research setting for the study of the influence of the ministry of finance for the function of evaluations. We argue that if similar patterns emerge across the two cases, then similar effects are also likely in other countries, where evaluation systems are anchored within the ministry of finance – regardless of the formality of such organisational anchorage.

Empirical data

To elucidate the constitutive power of the ministries of finance, we relied on a series of interviews among privileged informants, which we complemented with document analysis. For the interviews, we combined the perspectives from respondents situated at different positions within or independent of the central administration, each having a particular view on the countries' evaluation system from their respective roles. To allow for a systematic comparison, we targeted respondents with the same profile in both countries. We selected interviewees from four different positions. First, we interviewed civil servants working for the ministry of finance itself, and with a prominent position in advising and developing evaluation policies, and monitoring other ministries' evaluation activities. In the Netherlands, the Inspectorate of the Budget has an important role in this regard (Government Accounts Act, Article 34). The Bureau Strategic Analysis of this Inspectorate also took the lead in the government's large-scale Operation Insight into Quality (2018–2021), which came with a substantial revision of the Regulation for Periodic Evaluation in 2022. In the Danish case these tasks were not the formal responsibility of any specific section or division within the ministry of finance. Rather the tasks followed more informally from the work of two specific sections within the ministry of finance, from which we then collected our interviews. These were the sections responsible for developing, updating and coordinating the use of the methods and principles for calculating the economic effects of new policies, and the section responsible for coordinating the yearly budgetary process as well as overseeing the cost-calculations concerning new policy proposals. Second, we intentionally sought to compare the perspectives of civil servants in evaluation coordination positions within three central spending ministries: the ministries of health, the ministries of labour and the ministries of social affairs. Internationally speaking, the first two policy fields are among those with a strong institutionalisation of policy evaluations (see Stockmann et al, 2020), while the area of social policy is generally considered an area where evaluations are less systematised and institutionalised. We thus cover different types of ministries, as we know that evaluation cultures can differ across policy fields (Barbier and Hawkins, 2012; Jacob et al, 2015). The three spending ministries thereby function as exemplars of the consequences of the wider evaluation system, which we provide a more generic overview of by including interviews with evaluation experts from both courts of audits, who possess a comprehensive overview of the country's evaluation system.

This approach resulted in a total of 11 interviews involving 15 respondents from both countries. In Denmark, we conducted six interviews, involving a total of nine respondents from the ministry of finance (n=4), court of audit (n=2), ministry of interior and health (n=1), ministry of social affairs (n=1) and ministry of employment (n=1). In the Netherlands, we conducted five interviews, involving six respondents from the ministry of finance (n=2), court of audit (n=1), ministry of health, welfare and sports (n=2) and the ministry of social affairs and employment (n=1). The interviews, each lasting approximately one hour on average, were conducted between March and June 2023. They followed a semi-structured format, consisting of four main sections. In these sections, we addressed: (1) the organisation of the evaluation function, its current situation and evolution over time; (2) the various components of the evaluation system, aligned with the dimensions described by Leeuw and Furubo (2008); (3) the implications of the evaluation system; and (4) the needs and potential for further institutionalisation. The topics covered were derived from the literature on evaluation systems. A broad topic list was sent to all respondents in advance to allow them time to prepare for the interviews. The semi-structured approach enabled us to delve deeper into specific areas when necessary, and to adjust the order of questions as appropriate. All interviews were transcribed verbatim using artificial intelligence (AI) tools and then double-checked and edited by the authors. The interview quotes were first categorised according to the initial topic list, but then followed a second more abductive coding process, where quotes where recategorised in a back and forth process between the prevalent themes emerging from the interviews and the interpretive lens afforded by the concept of constitutive effects. Through this process the three-part structure of the analysis slowly emerged. The topic list and full questionnaire can be found at the public repository Data Station Social Sciences and Humanities: https://doi.org/10.17026/SS/VERXYA.

We supplemented the insights collected from the interviews with document analysis. Documents fell into two categories: first, we examined the prevailing evaluation guidelines and policies in the two countries; second, we reviewed available audits and (commissioned) analyses of specific aspects of the countries' evaluation systems. All documents were found by systematically going through the available publications of the ministry of finances, court of audits and the relevant ministries, which – for all publications since 2000 – are accessible through the websites of these organisations. These documents are referenced throughout our analysis, and they have also been listed at the public repository Data Station Social Sciences and Humanities (https://doi.org/10.17026/SS/VERXYA). Overall, while the document analysis provided a systematic assessment of the formal power of the ministry of finance, the interviews allowed us to capture the influence of more informal norms, discourse and practices that come from assigning a strong role to the ministry of finance in steering a country's evaluation system.

The constitutive effects of the role of the ministry of finance

We now turn to the consequences of the central role within the evaluation systems afforded to the ministry of finance, as we elucidate how it has constituted: (1) certain dominant evaluation methods and questions applied within central administration; (2) a specific primary user of the evaluation findings; and, following from this, (3) furthered accountability as the primary function of evaluations within central administration.

Evaluation criteria steeped in the language of accountants

Despite steps taken to further the epistemological coherence of the Dutch and Danish evaluation systems, both cases show a remarkable diversity in the types of evaluation methods and criteria being applied across individual ministries, as neither country has a formalised cross-ministerial hierarchy of evidence. However, there is evidence of more informal evidence hierarchies at play within both countries' central administration.

In Denmark, this hierarchy places (quasi)experimental methods – preferably randomised controlled trials – combined with cost–benefit calculations at the top and qualitative methods at the bottom. In practice, many other evaluation methods are applied across and within different ministries, but for evaluations to 'count' in the ministry of finance's macro-economic projections, these are the methods to use. The methodological hierarchy thus functions as a standard to aspire to, but it is not all ministries that are equally close to achieve it (Kelstrup and Jørgensen, 2024). When asked about which ministries have the strongest and most systematic approach to policy evaluations, all our respondents from the Danish ministry of finance as well as the respondents from the Danish National Audit highlight the same three ministries: 'The ministry of taxes, the ministry of employment and the ministry of education and research.' It is telling that these three ministries are also the exact same ministries where the ministry of finance most often calculate and include dynamic effects on the labour supply of the policies (Finansministeriet, 2018: 129–30; Kelstrup, 2024). This is further backed by our interviews with Danish respondents from the ministry of social affairs, ministry of health and ministry of employment. The ministry of employment is the only one of these ministries which has hitherto been allowed to include the expected dynamic effects of their policy proposals into the financing of said proposal. But to do so, this has entailed a very close collaboration with the ministry of finance. As explained by a respondent from the ministry of employment:

> It is very important that we calculate effects the same way [across the ministry of finance and the ministry of employment], because it has consequences for the economy. … This means that when there is a disagreement, we talk it over with the ministry of finance and land on an agreement on how to do the calculations.

The ministry of finance thus plays a very active role in guiding and judging the evaluations conducted within the area of employment policies, as these evaluations can potentially drive up spending, if they are able to document positive effects on the labour supply. As explained by a manager within the ministry of finance, this necessitates the insertion of as high a bar as possible for what counts as evidence: 'It is paramount for the whole way that we control expenses and public finances that we only include the things for which we trust in the evidence.'

The ministry of social affairs, on the other hand, experiences a much more hands-off approach by the ministry of finance regarding their evaluation activity, as their evaluations are not deemed methodologically sound enough to be incorporated *ex ante* into the macro-economic projections on the labour supply. A respondent from the ministry of social affairs explained how this illustrates the way the ministry of finance primarily sees evaluations as usable for 'keeping spending down', rather than as ways of gaining knowledge of 'what we get from the money we DO spend'.

A similar tendency can be found in the Netherlands, where the ministry of finance sees the two main policy-evaluation criteria of effectiveness and efficiency as critical to understanding how public money is being spent. These criteria are also built in the definition of policy evaluation in the Regulation. Policy evaluation is defined as 'research that provides insight into (conditions for) policy effectiveness and efficiency' (Ministerie van Financiën, 2022: Article 1). Individual ministries can, in principle, consider additional foci, such as 'public values' or 'broad welfare', as long as evaluation studies also address effectiveness and efficiency. Nonetheless, societal impact is sometimes felt to come at the cost of financial aspects. To put it in the words of a respondent: 'Money makes the world go around, but if you truly consider societal effects, the ministry of finance simply falls short in some instances … the financial perspective will always prevail.' Another respondent indicated that the ministry too often acts as an accountant with an emphasis on 'wrong frugality' and cost-saving.

The ministry's conceptualisation of policy effectiveness (and likewise, efficiency) also comes with an emphasis on quantitative evaluations. It is defined as 'the degree to which the policy objectives are achieved thanks to the policy deployed and with the least possible undesired side effects' (Ministerie van Financiën, 2022: Article 1). As also noted by a consultant in an evaluation commissioned by the ministry of health, welfare and sports (Regioplan, 2022: 2), this definition leaves little room for other types of evaluation research, and is not easily compatible with a more qualitative lens. Despite the relatively stringent definition, qualitative methods are getting increasingly appreciated, and they secured a place in the evaluation toolbox that was recently established by the ministry of finance.[1] Yet, and as is also the case in Denmark, they are still predominantly seen as having lesser value than more quantitative effect-evaluations. As a respondent of the ministry of finance put it in relation to qualitative methods:

> Yes, the methodological landscape is evolving, but this does not come without disadvantages. For instance, in the ministry of health, welfare and sports they were conducting a range of pilots, for which they were browsing a bit and sitting a bit together in groups. … This is evaluation research that does not deal with effectiveness and efficiency. This does make me a little itchy though.

Additionally, policy reviews frequently continue to rely on evidence hierarchies inspired by the Maryland Scientific Methods Scale. Evaluations that do not align with the relatively strict, quantitatively oriented definition of evaluation stipulated in the Regulation are often considered of lower quality (SEO, 2018), and therefore receive less importance.

Again, this is not to say that there are no qualitative evaluations being conducted, but they are still often perceived as inferior in terms of quality, and less capable of providing insights into the attribution of policies to specific objectives, and their efficient implementation. As perceived by a respondent in one of the policy ministries:

> We often feel that we need to interpret everything quantitatively. But this is not always possible nor necessary. With qualitative data you can also already offer insights into effectiveness and efficiency. But this is often a cry in the dark.

Despite the formal autonomy given to individual ministries to define their evaluation needs and select what they consider the most appropriate evaluation methods for a specific evaluation question, the central role of the ministry of finance within both the Dutch and the Danish evaluation systems has constituted a clear and shared ideal for 'good' evaluative evidence. An ideal that closely aligns with what commonly counts as appropriate from an accountant's perspective. Notably, this ideal is furthered not only by formal regulations but also by more informal processes. In both cases, the organisational anchorage of evaluations within the ministry of finance seeps into the language of evaluations which increasingly comes to resemble that of an accountant. Phrases such as 'balancing the books' and 'budgetary sobriety' are thus dominating discussion about evaluations within the central administrations of both countries and there is a tendency to prioritise effectiveness and efficiency as the central and essential evaluation criteria.

The ministry of finance becoming primary evaluation user

In the Netherlands it is stipulated in the regulatory framework (Ministerie van Financiën, 2022: Article 4) that the research design and report of periodic reviews are to be sent to Members of Parliament. In addition, though not mentioned in the Regulation, evaluations are also meant to inform policy makers, whether ministers or civil servants in charge of a particular policy. In Denmark, the parliament plays a lesser role due to the general *ex ante* focus of policy evaluations. Instead, the central administration typically produces evaluations to be used in processes of policy making by the relevant ministers and the civil servants within the relevant ministries. However, our findings suggest that in both cases, the evaluation system is instating the ministry of finance – and especially the top civil servants herein – as 'primus inter pares' when it comes to evaluation practice and thus the de facto primary user of the evaluations produced by the different ministries.

In the Netherlands, this role of 'primus inter pares' seems to directly follow from the formal responsibility given to the ministry of finance of coordinating the evaluation activities of the different ministries. While this in no way places the ministry of finance in the formal role of evaluation user, it gives the ministry substantial leverage to push for changes and incentivise other ministries to take evaluation practices seriously, 'whether they want it or not'. As explained to us by one of the respondents, different departments operate at different evaluation speeds, in different contextual circumstances, and have different backgrounds. This means that some departments will be frontrunners, while others will struggle to meet the requirements that come with the evaluation system and will lag behind. Considering these differences, the role of the ministry of finance is ultimately to keep everyone on track, which is not perceived possible with 'goodwill and support only'. Interestingly, however, as noted by the respondent of the court of audit, much leeway is possible for each ministry, as we were told, if they 'dare to engage in a conversation with finance' and provide a good proposal, justifying why deviation from default evaluation practice is desirable. Apparently, though, many departments have a tendency not to verify what is possible and sometimes act on the basis of misunderstandings. This behaviour can undeniably be explained by the mode of control with which the ministry of finance is traditionally associated and which is connected to their general role of controlling and overseeing the state budget. This role is mirrored in the requirement to submit the Strategic Evaluation Agenda along with ministries' budget. It is also exemplified by the key role played by the Directorates for Financial Economic Affairs (FEZ) in the management of evaluations in individual ministries (Ministerie van Financiën, 2020). It is most telling, for instance, that an evaluation coordinator in one of the policy departments consistently introduced herself to new policy colleagues as: 'I am X working for FEZ, but I am not a real FEZ-style person' with the intention to make them feel more at ease. Evaluations are thus still often perceived by ministries as obligatory and required by FEZ, who in turn refer to the ministry of finance. One respondent directly compared it with filling out taxes, where the purpose of evaluations risks becoming a 'box-ticking' exercise to accommodate the control and budgetary needs of the ministry of finance.

Since the introduction of the Strategic Evaluation Agenda in 2021, the link with the budgetary process has officially become less strict as ministries can now organise their evaluations by (policy) themes (Ministerie van Financiën, 2022: Article 2), as long as all articles (including the resources that fall under these articles) in the budget are still included in this agenda. According to some respondents, this helped to 'finally have a normal conversation' with the ministry of finance related to perceived knowledge needs, rather than about the policies covered by a particular budgetary article. At the same time, though, another respondent remarked that it is still primarily an 'accountability agenda', and that discretion is perceived as relatively limited, also when it comes to the terms by which policy themes need to be assessed (see also Berenschot, 2022: 35). As it was described to us, departments are on the defence and depend on the whim of the ministry of finance. Another respondent highlighted the use of 'war jargon' in the ministry (for example, 'shooting things in the agenda'), which was felt as symptomatic of the prevailing attitude.

In Denmark, we also find that evaluations are often both initiated and formed according to the needs and wishes of the ministry of finance. And here it is also a consequence of the social relations between spending ministries and the ministry of finance constituted by the tight coupling of the evaluation system with the budgetary process of the state. From the 1990s and onwards a norm has been developed, where all policy proposals by spending ministries are to be approved by the economic committee (Økonomiudvalget) before being presented to parliament. The ministry of finance functions as permanent chair of the economic committee, while the other ministries represented (typically four to eight) can change between governments. The permanence and power of this committee has over time become so institutionalised that it is often described as an informal cabinet of ministers, where the ministry of finance functions as de facto head of the cabinet. A respondent within the ministry of finance explains:

> Our power is in the coordination. … So the role of the ministry of finance in Denmark is different to most other countries, because our main task is to coordinate government policy. It is not because anybody has officially decided this, but it [the coordinating role] is just something that has been cultivated since the 1980s and which has become institutionalised.

In its role as permanent chair of the economic committee, the ministry of finance not only functions as gatekeeper for policy proposals, it is also increasingly taking on the role of gatekeeper for policy evaluations. Meaning that the ministry of finance validates and double-checks the calculations, evaluations and projections made by the spending ministry responsible for a given policy proposal presented before the economic committee. And just as in the Dutch case, this renders the ministry of finance the de facto primary user of policy evaluations – at least in the areas of tax, education and employment policies, where dynamic effects of policies are potentially included in the economic projections. In other welfare areas, such as health and social affairs, the ministry of finance is not as dominant as end user of the evaluation, as the potential effects of policies in those areas are currently deemed untranslatable into effects on the labour supply. As explained by a respondent from the ministry of health: 'We simply do not have an apparatus or knowledgebase, which is strong enough to do so [calculate the dynamic effects of policies]. We would like to do it, but we are simply not there yet.'

An evaluation system supporting the function of internal accountability rather than learning

Taken together, the furthering of an accountant's perspective on evaluations and the ministry of finance's role as de facto primary user of evaluations also furthers accountability and impedes learning as the primary function of both the Dutch and the Danish evaluation systems. In our cases the use of evaluations for learning is impeded in at least two distinct ways. First, by the way of routinisation, where the act of evaluating – and doing so in a certain way – becomes a goal in itself. Second, by fear of sanctions, which can impede the willingness of ministries to experiment and pose novel questions.

In the Netherlands, evaluations have often become 'box-ticking' exercises to accommodate the wishes of the ministry of finance. One of the respondents, for example, bemoaned the fact that the use of evaluations is sometimes compromised from the beginning when they are seen as obligatory, regardless of their quality. Initiatives to alleviate this problem have, however, been taken in recent years. With the introduction of the earlier mentioned Operation Insight into Quality, a range of initiatives have been taken to foster learning from evaluations, of which the pilot 'learning-oriented evaluations' of the ministry of health, welfare and sports is a case in point. The increasing attention to learning was confirmed and recognised by all respondents, also beyond this ministry, and has been said to result in growing attention for evaluation methods that are more participatory and that have learning as a key goal (for example, Folkert et al, 2018).

Still, the second impediment to learning continues to arise, as a consequence of the central position of the ministry of finance within the evaluation system, which furthers 'a mentality about money and not about policy'. Spending ministries thus fear that the results of their evaluations may have financial implications and threaten their policy portfolio, which is why several respondents describe how the civil servants of those ministries are not always fully 'honest' in their evaluations. The learning potential of evaluations thus to a large degree depends on the extent to which ministries 'dare' to evaluate, we were told, and especially dare to engage in a dialogue about the evaluation results. This fear of sanctions is thus impeding the willingness to openly discuss and question the desirability or appropriateness of a given policy, as also indicated by the court of audit in their assessment of the Operation Insight into Quality (Algemene Rekenkamer, 2021). Even when evaluations are used to improve policies, most respondents were under the impression that they primarily serve the purpose of 'tweaking' and fine-tuning existing policies, rather than leading to more substantial changes and rearrangements.

In Denmark, the ministry of finance is much more preoccupied with *ex ante* spending than *ex post* outcomes. The ministry of finance is thus highly focused on the process leading up to the passing of a policy reform and especially on double-checking and validating the projections concerning the potential economic effects of a given policy proposal. The same is not the case when a policy has been passed, as the ministry of finance – in the words of one of the respondents from within the ministry – is 'very reluctant on re-calculating and re-evaluating policies afterwards'. Another respondent from the ministry of finance explains that this focus on *ex ante* evaluations is 'super important as we are the ones responsible for ensuring that the right numbers are brought to bear in the fight for finances [between the ministries]'. This quite naturally limits the possibility to learn about the consequences of policies after they have been implemented. Second, the individual ministries also, like in the Dutch case, show a tendency for risk-avoidance, but in the Danish case it has less to do with the fear of being punished by the ministry of finance and more to do with a fear of 'missing out' on potential gains. This is especially vivid in the case of the Danish employment ministry, which has succeeded in producing high-ranked evaluations on certain aspects of employment policies. According to the respondent from the ministry of employment, this naturally skews the attention of policy makers in the direction of the few tried and tested programmes as 'there is a tendency that policy proposals flows in the direction of where there is already evidence or where we can calculate this [the dynamic effects]'. This tendency is also furthered by the difficulties in evaluating certain aspects of employment policies according to the high methodological standards of the evidence hierarchy. According to the respondent from the ministry of employment 'it is easy to gain evidence on the effects of caseworker-client meetings, as it is easy to measure whether you get a meeting or not'. Contrary to this, the respondent argues that more holistic services involving both employment, health and social services are 'much harder to measure and evaluate'. The economic gains of applying the tried and tested policy measures and the methodological difficulties in developing high-ranked evidence on the more complex aspects of employment policies thus limits the potential of the evaluation system to support the function of learning.

Discussion and conclusion

The analysis of the Dutch and Danish evaluation systems highlights the implications of assigning the evaluation coordination function to ministries of finance in terms of the intended function it serves. The function of evaluations are shaped not only – or primarily – by the ministry of finance forcing or coercing other ministries to adopt certain practices or procedures of evaluation, but also through processes where the logics, language and practices of the ministry of finance come to influence the evaluation activities of the whole central administration. In both countries, the result is that the ministry of finance's central role in the evaluation systems favours somewhat defensive qualities, where evaluations are primarily used for accountability and piecemeal changes in policies, rather than more fundamental learning and revisions or reflections on the appropriateness of specific policies – especially if such change includes a substantial increase in government spending.

These findings add to existing research in several ways. First, we elucidate some of the central mechanisms behind the influence of economists in general (Christensen, 2017; Maesse et al, 2022) and ministries of finance more specifically. Much of the existing literature has centred on documenting how the interests and logics of economists and the ministry of finance affect the content of actual policies (Hirschman and Berman, 2014; Berman, 2022; Garritzmann and Siderius, 2024). Less attention has been paid to the 'administrative underpinnings of expert influence' (Christensen, 2021). That is, how the organisation of knowledge and expertise within the procedures, structures and norms of policy making affects the ability of economists to assert their influence. By shifting focus from the influence of economists and economic knowledge on policy processes and output to the organisational factors supporting (or impeding) such influence of economists, our study can thus help explain the enduring influence of economic thinking even when the context and actors of policy-making processes change. Our findings thus mirror recent studies, which shows how neoliberal ideas continue to exert their influence on government policies, exactly because of the way these ideas have been inscribed into economic models or fiscal rules that continue to guide policy making (Maron, 2021; Fuglsang, 2023). Furthermore, our study provides new perspectives on the way the influence of economist thinking can extend beyond economic policies, rules and models, as the ministry of finance – through their central role in the evaluation system – helped further a specific epistemological and methodological position on knowledge production and use within the whole of the central administration.

Second, the article builds on the increasing focus of evaluation research on organisational factors influencing the uptake and use of evaluations (Kupiec et al, 2023a). Our study finds the institutional anchorage and organisation of evaluations to be of significant importance, but furthermore adds to this organisational perspective on evaluations by highlighting the importance of specific actors in shaping evaluation systems and their consequences. The consequences of the Danish and Dutch evaluation systems are not only a matter of the degree of institutional anchorage of evaluation activities within the central administration, but also a matter of *whom* the evaluation systems is centred around. Our two cases differed on the level of formalisation of the organisational underpinnings of the evaluation system, but this appeared to make less of a difference than the special role and standing that the ministry of finance holds. This suggest that the constitutive power of the evaluation systems is very much dependent on the central actors of such systems and would thus probably also be less vivid or different if the institutional anchorage of said systems was to be placed within a different ministry.

Third, the article contributes to the evaluation utilisation literature and especially debates on the ability of evaluations to be used for both learning and accountability (Alkin and King, 2017; Reinertsen et al, 2022). We thus concur with Raimondo (2018) that evaluation systems can lay out a blueprint and a set of norms favouring certain types of functions rather than others. And this blueprint can make it extremely challenging, if not impossible, to transform evaluation on different terms. Our study suggests that the very act of placing the responsibility of evaluations primarily in the hands of a powerful organisation and then planning evaluative activities according to the timeframes and knowledge needs of said organisation, can prove detrimental to the possibilities of learning. Meaning that the organising and planning of knowledge production to accommodate the existing frameworks and mental models of the organisation in charge will often contradict the ability of such evaluations to challenge underlying assumptions and thus induce more substantial changes. This is not due to ill-will within the ministry of finance, as all our respondents from the ministry highlight the need for broadening the perspective and functions.

However, such a broadening of the perspective and functions of evaluations will, on the one hand, require a greater awareness by the ministry of finance of the way their powerful position within central administration in general can also seep into the evaluation system and thus limit the methodological scope applied by other ministries. On the other hand, our findings also suggest that it requires the spending ministries to 'dare' to engage in dialogue with the ministry of finance, as this can help clarify how broader perspectives on evaluations are often possible.

That being said, there remain significant differences across individual departments, confirming the existence of sectoral evaluation patterns (De Francesco and Pattyn, 2021). Therefore, one should exercise caution in overinterpreting the findings. While we illustrate the major trends stemming from the institutionalisation of the evaluation system and the constitutive power of the ministry of finance, it would be erroneous to conclude that there is a uniform evaluation style. The choice of specific evaluation methods will also depend on sectoral evaluation styles, for instance, and the stage of the policy cycle during which evaluations are conducted. Additionally, individual ministries may exceed the (formal or informal) expectations set by the ministry of finance, and engage in additional evaluation practices that are not strictly required and where the ministry of finance has no role. Our analysis thus covers only a share of all evaluations conducted in both countries, though it represents the most significant share. Finally, the special nature of the two cases – regarding the central role of the ministries of finance and the long-standing emphasis and institutionalisation of procedures for using evaluative evidence – entails that we should not overemphasise the ability to generalise from these cases. Our findings indicate that formal procedures for anchoring evaluations within the ministry of finance probably means less than the general power and standing of the ministry of finance within the government hierarchy, which was substantial in both cases. This calls for more comparative and more contextually sensitive research on evaluation systems. Future research could thus ideally test to what extent our findings also hold in countries, where the ministry of finance does not have the same long-standing tradition and position as the main coordinator and arbiter of government policy. In addition, we also call for similar analyses of evaluation systems headed by other ministries in order to enhance our understanding of the influence of specific types of (economic) actors and organisational factors.

Note

[1] www.toolboxbeleidsevaluaties.nl.

Funding

Valérie Pattyn acknowledges funding of the Dutch Research Council (NWO), under grant number VI.Veni.211R.060.

Acknowledgements

We are grateful to the editor and the anonymous reviewers for their valuable comments and insights.

Contributor statement

NAA and VP jointly developed the idea for this article and contributed in equal parts to its conception, data collection, analysis and writing.

Research ethics statement

According to Danish legislation, approval from the ethics committee is not required for this type of study. For the data collected in the Netherlands, the research received the approval of the ethics committee of the Faculty of Governance and Global Affairs of Leiden University.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Algemene Rekenkamer (2021) *Inzicht in Kwaliteit. Operatie Geslaagd?*, Den Haag: Algemene Rekenkamer.

Alkin, M.C. and King, J.A. (2017) Definitions of evaluation use and misuse, evaluation influence, and factors affecting use, *American Journal of Evaluation*, 38(3): 434–50. doi: 10.1177/1098214017717015

Andersen, N.A. (2020) The constitutive effects of evaluation systems: lessons from the policymaking process of Danish Active Labour Market Policies, *Evaluation*, 26(3): 257–74. doi: 10.1177/1356389019876661

Barbier, J.C. and Hawkins, P. (2012) Conclusion: universal, common, diverse, and certainly not uniform, evaluation culture, in J.C. Barbier and P. Hawkins (eds) *Evaluation Cultures*, London: Routledge, pp 239–46.

Berenschot (2022) Eerste ervaringen met de strategische evaluatie agenda, Onderzoeksbijlage, https://open.overheid.nl/documenten/ronl-d1862a8b093c085da400a866bdfc2f2ad15a8617/pdf.

Berman, E.P. (2022) *Thinking Like an Economist: How Efficiency Replaced Equality in U.S. Public Policy*, Princeton: Princeton University Press. doi: 10.1515/9780691226606

Berman, E.P. and Hirschman, D. (2018) The sociology of quantification: where are we now?, *Contemporary Sociology*, 47(3): 257–66. doi: 10.1177/0094306118767649

Boswell, C. (2009) *The Political Uses of Expert Knowledge: Immigration Policy and Social Research*, Cambridge: Cambridge University Press.

Bourgeois, I. and Cousins, J.B. (2013) Understanding dimensions of organizational evaluation capacity, *American Journal of Evaluation*, 34(3): 299–319. doi: 10.1177/1098214013477235

Capano, G., Zito, A.R., Toth, F. and Rayner, J. (2022) Introduction: the state and public policy after the neoliberal wave – instruments and governance dynamics, in *Trajectories of Governance: How States Shaped Policy Sectors in the Neoliberal Age*, Cham: Springer International Publishing, pp 1–43.

Carlbaum, S. (2016) Customers, partners and rights-holders: school evaluations on websites, *Education Inquiry*, 7(3): 327. doi: 10.3402/edui.v7.29971

Christensen, J. (2017) *The Power of Economists Within the State*, Redwood City, CA: Stanford University Press.

Christensen, J. (2018) Economic knowledge and the scientization of policy advice, *Policy Sciences*, 51(3): 291–311. doi: 10.1007/s11077-018-9316-6

Christensen, J. (2021) Expert knowledge and policymaking: a multi-disciplinary research agenda, *Policy and Politics*, 49(3): 455–71. doi: 10.1332/030557320x15898190680037

Christensen, J. and Mandelkern, R. (2022) The technocratic tendencies of economists in government bureaucracy, *Governance (Oxford)*, 35(1): 233–57. doi: 10.1111/gove.12578

Dahler-Larsen, P. (2012) Constitutive effects as a social accomplishment: a qualitative study of the political in testing, *Education Inquiry*, 3(2): 171–86. doi: 10.3402/edui.v3i2.22026

Dahler-Larsen, P. (2014) Constitutive effects of performance indicators: getting beyond unintended consequences, *Public Management Review*, 16(7): 969–86. doi: 10.1080/14719037.2013.770058

Dahler-Larsen, P. (2024) The constitutive effects of measuring governance, in P. Triantafillou and J.M. Lewis (eds) *Handbook on Measuring Governance*, Cheltenham: Edward Elgar Publishing, pp 62–79.
De Francesco, F. and Pattyn, V. (2021) Policy evaluation styles, in M. Howlett and J. Tosun (eds) *The Routledge Handbook of Policy Styles*, New York: Routledge, pp 408–21.
De Francesco, F. and Pattyn, V.E. (2022) Designing for coordination: the case of regulatory management policy, in B.G. Peters and G. Fontaine (eds) *Research Handbook of Policy Design*, Cheltenham: Edward Elgar, pp 338–50.
Ejersbo, N., Greve, C., Ejersbo, N., Greve, C. and Ejersbo, N. (2014) *Moderniseringen af den Offentlige Sektor*, 3rd edn, Copenhagen: Akademisk Forlag Business.
Espeland, W.N. and Sauder, M. (2007) Rankings and reactivity: how public measures recreate social worlds, *The American Journal of Sociology*, 113(1): 1–40. doi: 10.1086/517897
Filgueiras, F. and Queiroz, L.D.F.N. (2021) The governance of public policy evaluation systems: policy effectiveness and accountability, *Organizações & Sociedade*, 28: 208–32.
Finansministeriet (2018) *Regneprincipper og Modelanvendelse – Dynamiske Effekter af Offentligt Forbrug og Offentlige Investeringer*, Copenhagen: Finansministeriet.
Finansministeriet (2021) *Økonomisk Analyse – Regneprincipper på Beskæftigelses- og Overførselsområdet*, Copenhagen: Finansministeriet.
Folkert, R., Verwoerd, L. and Verwest, F. (2018) Lerend evalueren: navigeren tussen verantwoorden en leren, *Beleidsonderzoek Online*. doi: 10.5553/BO/221335502018000007001
Foucault, M. (1990) *The History of Sexuality – Volume 1: The Will to Knowledge*, London: Penguin Books.
Fuglsang, N. (2023) The 'strange non-death' of economic models: how modelling contributed to neoliberal resilience in Denmark, *New Political Economy*, 28(5): 731–43. doi: 10.1080/13563467.2023.2172147
Furubo, J.E., Rist, R. and Sandahl, R. (2002) *International Evaluation Atlas*, New Brunswick: Transaction Publishers.
Garritzmann, J.L. and Siderius, K. (2024) Introducing 'ministerial politics': analyzing the role and crucial redistributive impact of individual ministries in policy-making, *Governance*, 1–22. doi: 10.1111/gove.12859
Hallerberg, M., Strauch, R. and von Hagen, J. (2006) *The Design of Fiscal Rules and Forms of Governance in European Union Countries, SFB/TR 15 Discussion Paper, No. 150*, München: Sonderforschungsbereich/Transregio 15 – Governance and the Efficiency of Economic Systems (GESY). doi: 10.5282/ubm/epub.13401
Hanberger, A. (2016) Evaluation in local school governance: a framework of analysis, *Education Inquiry*, 7(2): 217–36.
Hanberger, A. and Lindgren, L. (2019) Evaluation systems in local eldercare governance, *Journal of Social Work*, 19(2): 233–52. doi: 10.1177/1468017318760788
Haugaard, M. (2010) Power: a 'family resemblance' concept, *European Journal of Cultural Studies*, 13(4): 419–38. doi: 10.1177/1367549410377152
Heimberger, P., Huber, J. and Kapeller, J. (2020) The power of economic models: the case of the EU's fiscal regulation framework, *Socio-Economic Review*, 18(2): 337–66. doi: 10.1093/ser/mwz052
Hirschman, D. and Berman, E.P. (2014) Do economists make policies? On the political effects of economics, *Socio-Economic Review*, 12(4): 779–811. doi: 10.1093/ser/mwu017

Højlund, S. (2014) Evaluation use in evaluation systems: the case of the European Commission, *Evaluation*, 20(4): 428–46. doi: 10.1177/1356389014550562

Hood, C. (1995) The 'new public management' in the 1980s: variations on a theme, *Accounting, Organizations and Society*, 20(2–3): 93–109. doi: 10.1016/0361-3682(93)e0001-w

Hopwood, A.G. and Miller, P. (eds) (1994) *Accounting as Social and Institutional Practice*, Cambridge: Cambridge University Press.

Hult, A. and Edström, C. (2016) Teacher ambivalence towards school evaluation: promoting and ruining teacher professionalism, *Education Inquiry*, 7(3): 305. doi: 10.3402/edui.v7.30200

Jacob, S. and Varone, F. (2004) Cheminement institutionnel de l'évaluation des politiques publiques en France, en Suisse et aux Pays-Bas (1970–2003), *Politiques et Management Public*, 22(2): 135–52. doi: 10.3406/pomap.2004.2844

Jacob, S., Speer, S. and Furubo, J.E. (2015) The institutionalization of evaluation matters: updating the International Atlas of Evaluation 10 years later, *Evaluation*, 21(1): 6–31. doi: 10.1177/1356389014564248

Kaufmann, D. and Kraay, A. (2023) Worldwide governance indicators, 2023 update, www.govindicators.org.

Kelstrup, J.D. (2024) *Evidence-Based Policymaking and Public Administration in Denmark*, 1st edn, Cham: Palgrave Macmillan.

Kelstrup, J.D. and Jørgensen, J.V. (2024) Explaining differences in research utilization in evidence-based government ministries, *Policy Sciences*, 57: 257–80. doi: 10.1007/s11077-024-09529-6

King, J.A. and Alkin, M.C. (2018) The centrality of use: theories of evaluation use and influence and thoughts on the first 50 years of use research, *The American Journal of Evaluation*, 40(3): 431–58. doi: 10.1177/1098214018796328

Kok, J., Leistikow, I. and Bal, R. (2019) Pedagogy of regulation: strategies and instruments to supervise learning from adverse events, *Regulation & Governance*, 13(4): 470–87. doi: 10.1111/rego.12242

Kristiansen, M.B. (2018) Budget reforms in times of Austerity: a centralization cascade in Danish central government?, *Public Budgeting & Finance*, 38(3): 58–75. doi: 10.1111/pbaf.12191

Kupiec, T., Celińska-Janowicz, D. and Pattyn, V. (2023a) Understanding evaluation use from an organisational perspective: a review of the literature and a research agenda, *Evaluation*, 29(3): 338–55. doi: 10.1177/13563890231185164

Kupiec, T., Wojtowicz, D. and Olejniczak, K. (2023b) Structures and functions of complex evaluation systems: comparison of six Central and Eastern European countries, *International Review of Administrative Sciences*, 89(1): 202–20. doi: 10.1177/00208523211026964

Lægreid, P. (2017) Nordic administrative traditions, in P. Nedergaard and A. Wivel (eds) *The Routledge Handbook of Scandinavian Politics*, New York: Routledge, pp 80–91.

Lamarque, D. (2023) Strengthening evaluation capacity in government: why institutionalisation matters and why it is not sufficient, *Evaluation*, 29(2): 161–7. doi: 10.1177/13563890231166981

Leeuw, F.L. and Furubo, J.E. (2008) Evaluation systems: what are they and why study them?, *Evaluation*, 14(2): 157–69. doi: 10.1177/1356389007087537

Lindgren, J. (2015) Learning, coping, adjusting: making school inspections work in Swedish schools, *Nordic Journal of Studies in Educational Policy*, 3: 30126. doi: 10.3402/nstep.v1.30126

Maesse, J., Pühringer, S., Rossier, T. and Benz, P. (2022) *Power and Influence of Economists: Contributions to the Social Studies of Economics*, New York: Routledge.

Maron, A. (2021) Austerity beyond crisis: economists and the institution of austere social spending for at-risk children in Israel, *Journal of Social Policy*, 50(1): 168–87. doi: 10.1017/s004727942000001x

Mennicken, A. and Espeland, W.N. (2019) What's new with numbers? sociological approaches to the study of quantification, *Annual Review of Sociology*, 45: 223–45. doi: 10.1146/annurev-soc-073117-041343

Ministerie van Financiën (2020) *Inventarisatie Financiële Functie*, Auditdienst Rijk, https://www.rijksoverheid.nl/documenten/rapporten/2020/03/19/inventarisatie-financiele-functie#:~:text=De%20Auditdienst%20Rijk%20heeft%20een,functie%20bij%20de%20verschillende%20departementen.

Ministerie van Financiën (2022) *Regeling Periodiek Evaluatieonderzoek*, Den Haag: Ministerie van Financiën.

OECD (2020) *Improving Governance with Policy Evaluation: Lessons From Country Experiences, OECD Public Governance Reviews*, Paris: OECD Publishing. doi: 10.1787/89b1577d-en

Oudenampsen, M. and Mellink, B. (2022) The roots of Dutch frugality: the role of public choice theory in Dutch budgetary policy, *Journal of European Public Policy*, 29(8): 1206–24. doi: 10.1080/13501763.2021.1936130

Peters, B.G. and Pierre, J. (eds) (2004) *The Politicization of the Civil Service in Comparative Perspective: A Quest for Control*, vol 7, London: Routledge

Pollitt, C. and Bouckaert, G. (2017) *Public Management Reform: A Comparative Analysis-into the Age of Austerity*, Oxford: Oxford University Press.

Power, M. (1997) *The Audit Society: Rituals of Verification*, Oxford: Oxford University Press.

Preskill, H. and Boyle, S. (2008) A multidisciplinary model of evaluation capacity building, *American Journal of Evaluation*, 29(4): 443–59. doi: 10.1177/1098214008324182

Raimondo, E. (2018) The power and dysfunctions of evaluation systems in international organizations, *Evaluation*, 24(1): 26–41. doi: 10.1177/1356389017749068

Raudla, R., Douglas, J.W., Randma-Liiv, T. and Savi, R. (2015) The impact of fiscal crisis on decision-making processes in European governments: dynamics of a centralization cascade, *Public Administration Review*, 75(6): 842–52. doi: 10.1111/puar.12381

Regioplan (2022) Leren en/of verantwoorden. Ex Durante evaluatie van de pilot lerend evalueren, in *Opdracht van het Ministerie van Volksgezondheid, Welzijn en Sport*, Amsterdam: Regioplan.

Reinertsen, H., Bjørkdahl, K. and McNeill, D. (2022) Accountability versus learning in aid evaluation: a practice-oriented exploration of persistent dilemmas, *Evaluation*, 28(3): 356–78. doi: 10.1177/13563890221100848

Rist, R.C. and Stame, N. (eds) (2006) *From Studies to Streams: Managing Evaluative Systems*, vol 12, Piscataway, NJ: Transaction Publishers.

Roberts, A. (2010) *The Logic of Discipline: Global Capitalism and the Architecture of Government*, Oxford and New York: Oxford University Press.

Sanderson, I. (2002) Evaluation, policy learning and evidence-based policy making, *Public Administration*, 80(1): 1–22. doi: 10.1111/1467-9299.00292

SEO (2018) Beleidsdoorlichtingen belicht. Studie uitgevoerd in opdracht van het Ministerie van Financiën, https://www.seo.nl/publicaties/beleidsdoorlichtingen-belicht/.

Stockmann, R., Meyer, W. and Taube, L. (2020) *The Institutionalisation of Evaluation in Europe*, London: Palgrave Macmillan.

van der Knaap, P., Pattyn, V. and Hanemaayer, D. (2020) *Beleidsevaluatie in Theorie en Praktijk*, The Hague: Boom Bestuurskunde.

Vedung, E. (2010) Four waves of evaluation diffusion, *Evaluation*, 16(3): 263–77. doi: 10.1177/1356389010372452

Vedung, E. (2017) *Public Policy and Program Evaluation*, 2nd edn, New York: Routledge.

Verhoest, K. (2016) The relevance of culture for NPM, in *The Ashgate Research Companion to New Public Management*, New York: Routledge, pp 47–63.

Weiss, C.H. (1998) Have we learned anything new about the use of evaluation?, *American Journal of Evaluation*, 19(1): 21–33. doi: 10.1177/109821409801900103