



## Accounting, Calculative Infrastructures and Commensuration Work

Fredrik Svärdsten & Sven Modell

**To cite this article:** Fredrik Svärdsten & Sven Modell (2023): Accounting, Calculative Infrastructures and Commensuration Work, European Accounting Review, DOI: [10.1080/09638180.2023.2166091](https://doi.org/10.1080/09638180.2023.2166091)

**To link to this article:** <https://doi.org/10.1080/09638180.2023.2166091>



Published online: 20 Jan 2023.



Submit your article to this journal [↗](#)



Article views: 241



View related articles [↗](#)



View Crossmark data [↗](#)

# Accounting, Calculative Infrastructures and Commensuration Work

FREDRIK SVÄRDSTEN\* and SVEN MODELL\*\*

\*Stockholm Business School, Stockholm University, \*\*Alliance Manchester Business School, University of Manchester NHH – Norwegian School of Economics and Turku School of Economics

(Received: 22 September 2021; accepted: 23 December 2022)

**ABSTRACT** Accounting scholars are paying increasing attention to the notion of commensuration, or the translation of different objects into common metrics that make the entities that are subject to performance evaluation comparable, and the effects this has on organizational behavior. This body of research draws attention to organizations' commensuration work, defined as the efforts that evaluated organizations expend in pursuing and dealing with resistance to commensuration, but has mainly explored how such work emerges in response to a single set of comparable metrics such as individual rankings or ratings. By contrast, we ask how commensuration work unfolds in relation to a broader assemblage of calculative infrastructures. Such infrastructures can entail multiple, inter-related performance evaluation procedures that influence organizations' commensuration work and give rise to complex patterns of reactivity, denoting the propensity of organizations to alter their behavior as they internalize external pressures for performance evaluation. Our empirical findings underline the importance of adopting a multi-layered view of commensuration work, that entails both direct and indirect forms of reactivity, as it unfolds in relation to a broader assemblage of calculative infrastructures. We discuss the implications of adopting such a view for future research on commensuration and related research on performance evaluation.

**Keywords:** Accounting; calculative infrastructure; commensuration; reactivity

## Introduction

Accounting scholars have long taken a keen interest in the regulatory standards that govern the development of organizational accounting practices as they evolve as an integral part of broader, societal reform programs (Miller, 1991; Modell, 2014). A particular type of standardization that is gaining increasing recognition in this body of scholarship is that concerned with the notion of commensuration, or the translation of different objects into common metrics that make the entities that are subject to performance evaluation comparable (Espeland & Stevens, 1998, 2008). Accounting research exploring this phenomenon has mainly evolved along two trajectories. The first stream of research has focused on how mechanisms of commensuration, such as the categorization of qualitative phenomena into comparable metrics, come into being and govern broader populations of organizations (e.g., Crvelin & Löhlein, 2022; Mehrpouya

---

*Correspondence Address:* Sven Modell, Alliance Manchester Business School, University of Manchester, NHH – Norwegian School of Economics and Turku School of Economics, Booth Street West, Manchester M15 6PB, UK. Email: [sven.modell@manchester.ac.uk](mailto:sven.modell@manchester.ac.uk)

Paper accepted by Beatriz Garcia Osma.

& Samiolo, 2016; Pollock & D'Adderio, 2012; Unerman et al., 2018). The second and more substantial body of research examines how evaluated organizations respond to such mechanisms (e.g., Englund & Gerdin, 2020; Gerdin & Englund, 2019; Kolk et al., 2008; Kornberger & Carter, 2010; Llewellyn & Northcott, 2005; Scott & Orlikowski, 2012). Much of this research documents how such responses are underpinned by strong tendencies toward reactivity, or the propensity of organizations to alter their behavior as they internalize external pressures for performance evaluation (Espeland & Sauder, 2007, 2016), but also draws attention to how such reactive tendencies may be contested by actors who resist or strive to transform the mechanisms through which commensuration is brought about (e.g., Englund & Gerdin, 2020; Gerdin & Englund, 2019; Mennicken, 2013; Pollock et al., 2018; Samiolo, 2012).

While prior research on commensuration provides valuable insights into what some authors call commensuration work (Habran & Mouritsen, 2022; Rindova et al., 2018), or the efforts that evaluated organizations expend in pursuing and dealing with resistance to commensuration, it has mostly examined the development and impact of a single set of metrics such as individual rankings or ratings.<sup>1</sup> With the exception of a small number of studies that examine how organizations respond to competing rankings or ratings (Dorn, 2019; Pollock et al., 2018; Sauder & Espeland, 2006; Slager & Gond, 2022), little attention has been paid to how commensuration evolves in response to a wider range of co-existing performance evaluation procedures. This underplays the potential complexity of commensuration work. Extending research on commensuration work to examine how it is influenced by a wider range of performance evaluation procedures, that are not necessarily in competition with each other, is not least important in contexts where organizations are governed by a broader assemblage of calculative infrastructures. Calculative infrastructures have been defined as a 'relatively stabilized chain of accounting calculations and associated narratives' (Kurunmäki & Miller, 2013, p. 1101) and may entail multiple, inter-related performance evaluation procedures that help to establish and maintain societal reform programs (Juven, 2019; Reilley & Scheytt, 2019).

The discussion above raises the question of how organizations' commensuration work unfolds where a broader assemblage of calculative infrastructures underpins societal reform programs. We address this research question in the context of Swedish central government during a period when calculative infrastructures, that entailed a considerable element of commensuration, were partly de-stabilized by governance reforms that sought to radically increase the reporting discretion of Swedish government agencies. We show how these reforms led an individual government agency to engage in a significant amount of internally initiated commensuration work to advance new performance reporting practices, but that this work was reinforced by a broader assemblage of calculative infrastructures that contributed to the stabilization of these practices. In contrast to prior research on commensuration, these insights lead us to advance a multi-layered view of commensuration work that recognizes that such work is not necessarily confined to efforts emerging as a direct reaction to pre-existing, comparable metrics, but that it may also entail indirect forms of reactivity that are implicated in a reciprocal interplay with commensuration. While the commensuration work initially emerged as a response to a lack of clear guidelines for how Swedish government agencies should structure their performance reporting, it was gradually connected to a broader assemblage of calculative infrastructures which reinforced indirect forms of reactivity that led to further commensuration work and ultimately constrained performance reporting. We extend this reciprocal, multi-layered view of commensuration work by

---

<sup>1</sup>Rankings and ratings differ insofar as the former involve the assignment of a unique, rank-ordered value to each evaluated entity while the latter can, in principle, assign the same value to multiple entities. However, both types of performance evaluation rely on a process of commensuration whereby evaluated entities are made comparable across particular objects of evaluation (see Ringel et al., 2021)

distinguishing between three types of work – categorization work, metricization work and tracing work – as a basis for explaining how such work evolves and discuss the implications for future research on commensuration.

The remainder of the paper is organized as follows. We start by outlining how evaluated organizations' commensuration work has been conceptualized in prior research and advance a preliminary theoretical framework for our analysis. We then introduce our research context and methods before describing how the commensuration work unfolded in our field study site. We conclude the paper with a discussion of our findings and their implications for future research.

## **Theoretical Framework**

The notion of commensuration occupies a central position in the sociology of quantification (Espeland & Stevens, 2008; Mennicken & Espeland, 2019) and has attracted considerable attention as the bedrock of performance evaluation practices that rely on the ability of decision-makers to make judgements about the relative performance of evaluated entities (see e.g., Espeland & Sauder, 2007; Esposito & Stark, 2019; Jeacle & Carter, 2011; Kolk et al., 2008; Mehrpouya & Samiolo, 2016; Stark, 2020). In contrast to modes of evaluation that treat the objects of evaluation as unique, uncertain and multi-dimensional singularities (Bialecki et al., 2017; Karpik, 2010), commensuration relies on the translation of different objects into common metrics that facilitate comparisons across evaluated entities (Espeland & Stevens, 1998, 2008). As such, commensuration imbues metrics with an air of certainty and accuracy, even though the underlying object of evaluation is variable and not always possible to standardize.

Following the view of commensurable metrics as imbued with an air of certainty and accuracy, it is perhaps not surprising that organizations' commensuration work is often seen as a response to their need for uncertainty reduction and legitimacy in the eyes of external audiences. According to Espeland (1998, p. 24), an important driving force behind organizations' pursuit of commensuration is the urge 'to impose order, exert control, or manage uncertainty'. Even though the uncertainty reduction that follows from commensuration is often incomplete (Esposito & Stark, 2019; Habran & Mouritsen, 2022; Sauder et al., 2021; Sauder & Espeland, 2009), the use of commensurable metrics may fill organizations with a sense of order that enhances their confidence that they are behaving appropriately. Similarly, compliance with performance metrics that make organizations comparable with their peers bestows social legitimacy on them because it enhances society's trust that they are performing in accordance with acceptable norms and standards (Espeland & Stevens, 1998; Jeacle & Carter, 2011; Porter, 1995). Such legitimacy can, in turn, reinforce organizations' confidence in the appropriateness of their actions and provides a basis for imbuing organizational behavior with a sense of order and predictability. Hence, as driving forces behind commensuration work, organizations' need for uncertainty reduction and legitimacy are often closely intertwined and not always possible to disentangle in empirical inquiries (Espeland & Stevens, 1998).

To unpack the notion of commensuration work, it is meaningful to disaggregate this construct into at least three distinct but inter-related types of work that play different roles in organizations. The first type of work is what we will refer to as categorization work. According to Espeland and Stevens (1998), commensuration typically starts with the creation of new categories, or novel objects of comparison, that become constitutive of social realities. Such categorization work is distinct from the efforts to make categories measurable or what we will refer to as metricization work. Such quantification efforts are a key part of the process of establishing the relative position of evaluated entities along the categories that are included in performance evaluation and imbue such entities with value (Espeland & Stevens, 1998; Ringel et al., 2021). Finally, we add a third type of work, labeled tracing work, to capture the efforts to connect commensuration to

calculative infrastructures. As noted by Power (2019, 2021), the existence of such infrastructures often implies that actors around and within organizations have an interest in tracing the chains of calculation that influence their accounting practices and making these connections visible for the sake of legitimizing such practices. Correspondingly, we may conceive of the efforts to connect a focal set of comparable metrics to a broader assemblage of calculative infrastructures within and beyond organizations as a particular type of work that is distinct from, but potentially influencing, their ability to create new categories and make such categories measurable.

While commensuration work, such as that outlined above, may emerge as a reactive response to external pressures that aims at reducing uncertainty and enhancing or maintaining organizational legitimacy (Espeland & Sauder, 2007, 2016), such work is often subject to contestation that weakens reactive tendencies. The likelihood of contestation is partly dependent on the extent to which diverse organizational actors harbor a pronounced commensuration mindset, defined as an inherent propensity to seek out and use comparable metrics, or whether they are inclined to champion other modes of evaluation that detract from such metrics (Espeland, 1998; Espeland & Stevens, 1998). The extent to which one or the other mindset dominates can, in turn, be understood against the backdrop of the interests and values that different actors espouse. Since commensuration is often an integral part of organizational valuation processes (Fourcade, 2011; Jeacle, 2022; Lamont, 2012), it nearly inevitably raises questions about whose values are represented by commensurable metrics and whose values are at risk of being marginalized by the use of such metrics. For instance, in her seminal study of commensuration within the US Bureau of Reclamation, Espeland (1998) documents how a group of bureaucrats, who displayed a pronounced commensuration mindset, engaged in forceful quantification efforts and how these efforts were contested by actors whose professional values were largely incompatible with the mode of valuation that followed in the wake of commensuration. Following Lamont (2012), we may conceive of such conflicts as rooted in heterarchies of values. The existence of such heterarchies means that a set of heterogeneous values co-exist without being subordinate or superordinate to each other and this, in turn, is likely to render commensuration an inherently contested phenomenon.

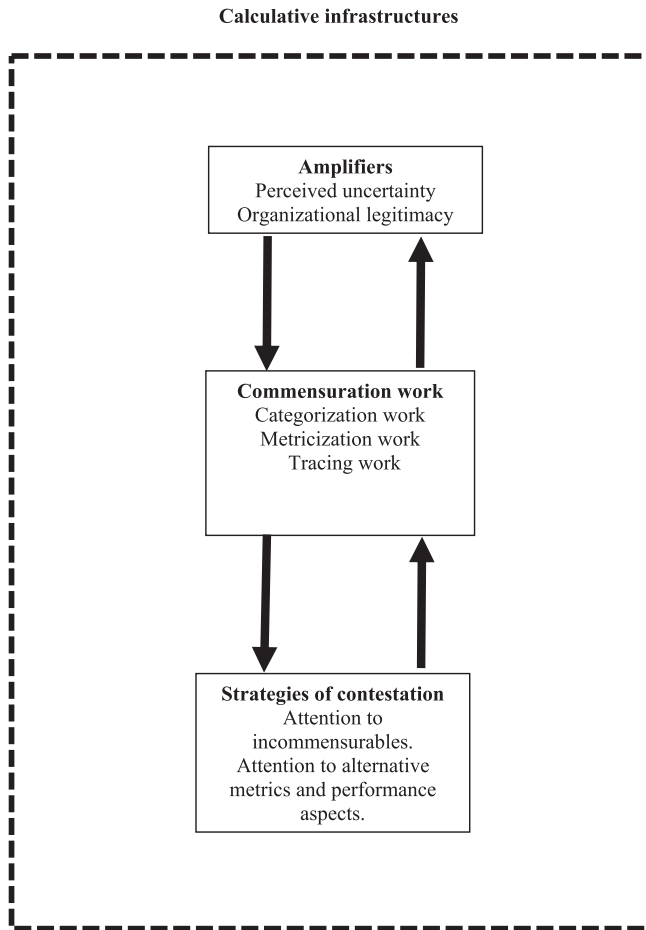
Accounting scholars have especially documented two strategies of contestation that often emerge in response to organizations' commensuration work but that can also intensify the work aimed at dealing with resistance. The first is for organizational actors who resist the use of supposedly commensurable metrics to draw attention to incommensurables, or the unique properties of the objects of evaluation that such metrics occlude (Englund & Gerdin, 2020; Espeland, 1998; Gerdin & Englund, 2019; Mennicken, 2013; Rainelli-Weiss & Huault, 2016; Samiolo, 2012). This strategy of contestation is most likely to occur where actors agree about which metrics should be used for evaluation, but disagree about how such metrics should be used. For instance, in her study of the use of cost-benefit analyses to make the environmental impacts of the flood defences of Venice commensurable, Samiolo (2012) found that actors who resisted specific calculations sought to problematize them by highlighting the unique and differentiated natural environment that they purported to reflect although they did not necessarily contest the need to calculate benefits and costs as such. Similarly, studies of commensuration in universities have found that different actors may agree about which metrics should be used for performance evaluation, but disagree about whether it is appropriate to use such metrics for comparing performance across different entities and thus obscuring the uniqueness of what is being evaluated (Englund & Gerdin, 2020; Gerdin & Englund, 2019). To deal with such resistance, organizations may need to engage in additional commensuration work that aims at alleviating conflicts. Hence, we need to conceive of commensuration work as an organizational phenomenon that is implicated in a reciprocal relationship with the attempts of some actors to draw attention to incommensurables.

The second, major strategy of contestation that has been documented in the literature on commensuration is for organizational actors to draw attention to alternative metrics or performance aspects to those being used for the purpose of commensuration in an attempt to challenge them or resist compliance with a single set of metrics (Dorn, 2019; Elsbach & Kramer, 1996; Huault & Rainelli-Weiss, 2011; Pollock et al., 2018; Sauder & Espeland, 2006, 2009; Slager & Gond, 2022). This strategy of contestation is different from attempts to draw attention to incommensurables in that it does not necessarily question the appropriateness of basing evaluations on commensurable metrics, but rather challenges which metrics should be used for this purpose. For instance, in their study of the influence of multiple rankings promoted by information technology analysts, Pollock et al. (2018) showed that the organizations that were being ranked did not slavishly comply with specific rankings but used diverse rankings, that focused on slightly different performance aspects, to challenge and even transform the metrics that were being used to rank them through their continuous dealings with such analysts. This suggests a view of commensuration as a malleable phenomenon, that is not necessarily imposed on organizations in a ready-made format, but that rather evolves as a result of a reciprocal interplay between the commensuration work conducted by propagators of diverse performance evaluation procedures and the attempts by evaluated entities to manipulate such procedures. Similar to the additional commensuration work that may be required to alleviate conflicts due to attempts to draw attention to incommensurables, it is thus useful to conceive of such work as a dynamic phenomenon that may both cause actors to mobilize alternative metrics or performance aspects and that gradually evolves as a response to this strategy of contestation.

The above discussion leads us to propose a theoretical framework for examining how commensuration work unfolds (see Figure 1). We posit that the root causes of such work, that amplify organizations' reactive responses to external pressures, are their need for uncertainty reduction and legitimacy. Insofar as organizations' commensuration work is successful and relatively uncontested, we may also expect the levels of perceived uncertainty to be reduced and organizational legitimacy to be enhanced or maintained. However, since commensuration work can be contested, it is also important to recognize how strategies of contestation, such as the attempts by organizational actors to draw attention to incommensurables and alternative metrics and performance aspects, emerge and affect the continuous unfolding of such work. While these strategies of contestation may emerge as a consequence of the initial commensuration work that is undertaken by organizations, they are also likely to be implicated in a reciprocal interplay with the work that is involved in dealing with resistance. In our empirical analysis, we extend this view of how commensuration work unfolds by examining how it relates to a broader assemblage of calculative infrastructures. In doing so, we treat the salience of such infrastructures as a general, theoretical boundary condition (as denoted by the dotted line in Figure 1) and then flesh out the exact role of such infrastructures in relation to the commensuration work under examination as a basis for further theory development.

## **Research Context and Methods**

Our empirical study is set in the context of Swedish central government. This context constitutes a relatively unique setting in the sense that government agencies, which constitute the executive branch of central government, have long enjoyed a high degree of discretion as part of a highly devolved system of governance. Several commentators have noted that Swedish government agencies are more sheltered from direct interventions by elected politicians in operating matters than similar organizations in many other jurisdictions (Modell, 2021; Molander et al., 2002; Pierre, 2004). The ministers in charge of the government offices to which the agencies are held accountable are constitutionally prohibited from intervening in such matters. This has



**Figure 1.** Theoretical framework.

enhanced the autonomy of Swedish government agencies and contributed to institutionalizing an ideal conception of them as relatively unconstrained insofar as the design of control practices is concerned.

Starting in the late 1980s, a partially new, calculative infrastructure that was, to a large extent, inspired by the diffusion of New Public Management discourses (see Hood, 1991, 1995) emerged within this system of governance. The starting point for this development was the introduction of management by objectives as a means of further deregulating governance practices and inducing a long-term planning horizon extending beyond the annual budgetary cycle (Modell et al., 2007; Sundström, 2003). Management by objectives came to rest on the agencies' appropriation directives, which set out their annual performance targets and specified the financial resources available to meet those targets while also governing the format of their annual reports to the Government. The reporting format was determined by the so-called activity structure in the appropriation directives which divided agency operations into different activity areas. The activity structure aimed at making especially the costs of activity areas comparable across the agencies in order to facilitate the Government's allocation of resources and also offered a common template for linking such costs to quantifiable performance indicators. As such, the appropriation directives entailed an element of commensuration even though the tasks entrusted to the agencies vary considerably and are rarely comparable in operational terms.

The introduction of management by objectives was followed by reforms that transformed other parts of the calculative infrastructures governing Swedish government agencies. In 1998, the National Audit Office (NAO) was split into two organizations to separate its advisory role from its auditing practices (Sundström, 2003). The advisory role was vested in the Swedish National Financial Management Agency (SNFMA), which functions as the Government's expert on accounting and control issues. Even though it has no mandatory regulatory powers, it is charged with the development and dissemination of 'good accounting practices' in central government and supports Swedish government agencies in interpreting regulations issued by the Government and Parliament. The norms and guidelines emerging from its work may be seen as the basis for calculative infrastructures that complement those provided by the NAO, which subsequently became a pure auditing body with the audits of the agencies' annual reports and more targeted performance audits focusing on specific operating areas as its main evaluation procedures (Grönlund et al., 2011). To some extent, the evaluation procedures of the NAO and the SNFMA also remained interlinked. Since its inception, the SNFMA has conducted the so-called Financial Management Evaluations, which are broadly based assessments of the accounting practices and related administrative procedures of Swedish government agencies, on an annual basis. Critical remarks from the NAO can have a negative impact on these evaluations, which entail a rating of most agencies along a number of comparable metrics reflecting the quality of their accounting practices. Hence, the Financial Management Evaluations imposed yet another element of commensuration on the agencies.

The reforms described above established a relatively stable assemblage of calculative infrastructures (cf. Kurunmäki & Miller, 2013; Power, 2015) that underpinned a societal reform program pivoting on the need to enhance agency autonomy and accountability for results. However, following growing criticisms that especially management by objectives failed to reduce the element of detailed control, but rather increased the agencies' reporting requirements, additional reforms that de-stabilized parts of this infrastructure emerged in the 2000s. In response to these criticisms, the Government initiated a comprehensive inquiry, called the Performance Management Investigation, in 2006 to assess the efficacy of management by objectives and provide recommendations for its future development (Swedish Government, 2007). The Performance Management Investigation proposed relatively far-reaching changes that were aimed at simplifying the agencies' performance reporting and, not least, enhancing their discretion in this regard (see Svärdsten, 2015).

To further agency autonomy, the Performance Management Investigation advised the Government to abolish the activity structure in the appropriation directives and replace the attempts to couple resource allocation decisions to comparable costs with the distribution of block grants which should then be re-allocated within the agencies. It also recommended that the agencies should effectively be allowed to determine their own, short-term performance targets and adapt their reporting to better reflect their unique, operating characteristics. This emphasis on the uniqueness of agency operations can be seen as an attempt to affirm their singularities (cf. Bialecki et al., 2017; Karpik, 2010) and reduce the emphasis on commensuration that was previously afforded by the appropriation directives. The only restrictions that were being proposed were that the performance targets and reporting should be based on measurable outputs and linked to the costs associated with delivering specific outputs and that their performance reporting should be structured in ways that enabled the Government to track their performance over time.

New regulations that largely reflected the recommendations of the Performance Management Investigation came into force in 2009 (Swedish Parliament, 2009). Around the same time, the first author of this paper initiated a field study in the Swedish Energy Agency (SEA) which extended between October 2009 and December 2011. The SEA was founded in 1998 and at the

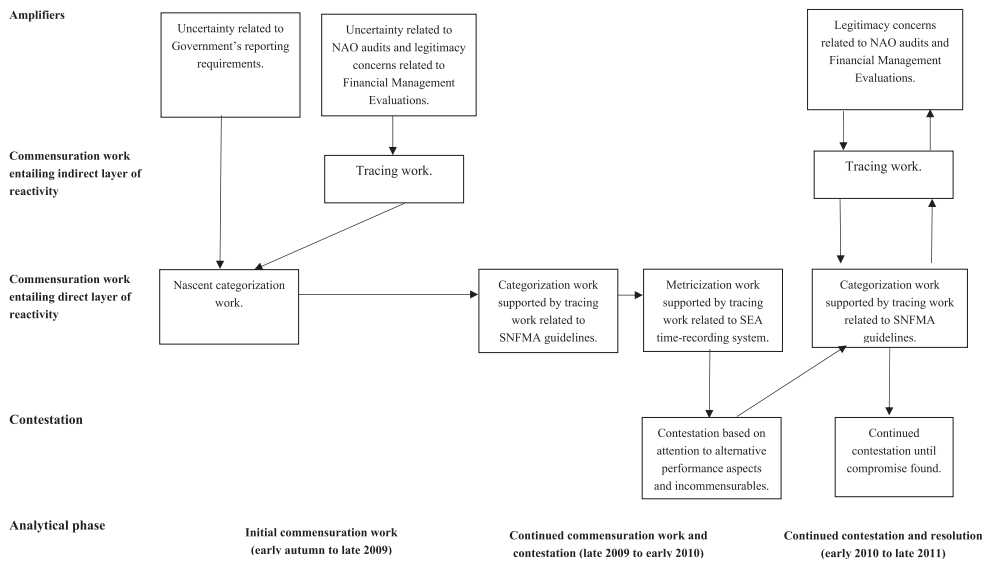


time of our study it had approximately 350 employees working in six departments. The agency's mission is to further the development of an economically and ecologically sustainable energy system at a national as well as global level and it has historically been dominated by professional values that strongly support this mission. The vast majority of its employees have an academic degree and many of them possess significant expertise in the areas of energy production and sustainability. The agency supports research for increased energy efficiency and provides both services, such as knowledge support for energy efficiency in municipal and industrial networks, and physical products, such as electronic energy efficiency calculation models for private households. It also acts as an expert on energy-related policy issues and supports the Government in relation to specific events, such as international conferences, and through ongoing consultations on energy-related matters. The SEA is governed by the Ministry of Enterprise, Energy and Communications (MEEC), with which it has a close working relationship, but also receives assignments from the Ministry of the Environment.

Our field study in the SEA followed the process of determining the format of the agency's annual report, which entailed a considerable element of commensuration work, as it evolved in real time as a response to the Government's changing reform agenda. All data collection was conducted by the first author and mainly comprised interviews, observations and consultation of various documentary sources. Consistent with ethical guidelines for academic research in Sweden, informed consent was obtained orally. The selection of interviewees was primarily based on the criterion that they should be involved in the process of determining the new reporting format and, especially, the definition of the agency's outputs that formed a key part of this process. Another criterion was that the interviewees must represent all departments within the agency, as we wanted to know whether various activities were more or less difficult to translate into the emerging reporting format. Subject to these criteria, a snowball sampling approach was adopted for selecting interviewees in the SEA (Berg & Lune, 2012).

In total, 20 interviews with 16 individuals were undertaken in the SEA. Key informants, such as the Chief Accountant and two members of the Central Office, were interviewed on several occasions. In addition to the interviews within the SEA, the first author interviewed three civil servants within the MEEC who have extensive contacts with the agency. He also interviewed and held extensive, informal discussions with two of the civil servants in charge of the Performance Management Investigation and a civil servant working for the SNFMA (see Online Appendix A). The interviews generally lasted between 60 and 90 min and followed a semi-structured format. With the respondents' permission, all interviews were digitally recorded and were complemented with manual notes to document emerging impressions and non-verbal cues. In conjunction with the interviews, the first author also had ample opportunities to engage informally with the interviewees.

The interviews were complemented with observations of four meetings that aimed at finding a way of modifying the definition of output in response to emerging attempts to contest the agency's commensuration work (see Online Appendix B and C). The meetings were documented through digital recordings and note-taking by the first author. The meeting observations played a valuable role in our analysis as they provided rich insights into the conflicts that emerged from the SEA's initial commensuration work and how actors promoting this work sought to overcome resistance. The analysis was further enriched by various types of documents, such as relevant government inquiries, bills, reports published by the SNFMA, the SEA's annual reports, operating plans, meeting protocols and internal working material and presentation material related to the process under examination. Our study was also informed by a parallel study of the evolving auditing practices of the NAO during the period under examination (see Grönlund et al., 2011; Svårdsten, 2019).



**Figure 2.** Summary of empirical analysis.

The initial analysis of the data took the form of ongoing, open-ended coding as the research unfolded and resulted in a relatively extensive report that accounted for key empirical findings. The first author had the opportunity of validating this analysis through presentations to selected staff within the SEA. Our more theoretically informed analysis reported in the present paper then followed a largely abductive approach (Lukka & Modell, 2010; Timmermans & Tavory, 2012) that aimed at advancing ‘thick explanations’ (Lukka & Modell, 2010, p. 466) of the commensuration work unfolding in the SEA. Advancing such explanations required us to ground our etic, theoretically informed analysis in a rich process description providing a context-sensitive, emic understanding of how this work evolved (Lukka & Modell, 2010, 2017). Through this analysis, we gradually came to realize that the commensuration work unfolding in the SEA could be divided into two overriding categories of work where one entailed a direct layer of reactivity, emanating from the Government’s changing reporting requirements, while the other entailed a more indirect layer of reactivity that was somewhat removed from but still influenced the agency’s performance reporting. We then connected this emerging, multi-layered view of commensuration work to the three types of work derived from prior research and teased out how these types of work influenced each other across three relatively distinct, analytical phases (see Figure 2). The following empirical narrative closely follows these analytical phases.

## The Commensuration of Performance Reporting in the Swedish Energy Agency

### *Initial Commensuration Work (Early Autumn to Late 2009)*

The new directives for performance reporting that were introduced in 2009 created considerable uncertainty among Swedish government agencies. The abolishment of the activity structure in the appropriation directives led to a lack of clarity as to what should be seen as outputs and how output metrics should be linked to costs (Svårdsten, 2015). These problems were especially palpable in the SEA, where the concept of output was relatively new. The agency’s performance reporting had historically followed its appropriation directives very closely and was divided into

six activity areas. Each activity area had a set of quantifiable performance indicators, but the indicators were not necessarily output-based and, in many cases, rather reflected underlying, operating activities. Also, the agency had little experience of systematically distinguishing such activities from outputs when constructing its annual reports and complemented its performance reporting with extensive narratives that enabled employees to describe their activities and their professional opinions about the effects on society. This tendency to produce lengthy narratives was arguably reinforced by the professional identities and values of the civil servants working for the SEA:

R9: Every year, the auditors [i.e., the NAO] say that the annual report is too exhaustive. We write and write and write, I often say we are a writing agency.

Interviewer: Why are you being so verbal?

R9: Look at who is working here. People come from academia [...] 80–85 per cent of them have an academic degree and 15 per cent have a PhD. Language is very much a tool for us. It's simply part of our culture to produce so much text. (Interview, 2010-03-24).

To comply with the emerging reform agenda and define the agency's outputs a working group, consisting of representatives from the agency's senior management team and all of its departments, was formed in the early autumn of 2009. The work was led by the Chief Accountant, who was ultimately responsible for the agency's performance reporting, and initially created considerable enthusiasm. The work on defining output was reportedly seen as a healthy learning process that, according to one interviewee, would clarify what the agency did and enhance its discretion as to what to account for:

R4: Earlier, there were frequent complaints [within the agency] that we were so restricted by the appropriation directives and that we were only allowed to account for what it stipulated rather than what we would like to report, but now we have said that we are open to account for 'all the other things' [...] The new [reporting] structure has created space for more open-ended thinking. (Interview, 2009-11-09).

This initial enthusiasm suggests that the Government's requirement for agencies to produce measurable output indicators was at least initially uncontested and not the root cause of the problems that subsequently accompanied the commensuration work in the SEA. The problems rather emerged as the agency engaged in more concerted categorization work, which initially exacerbated rather than reduced the uncertainty surrounding the definition of outputs. In the first meeting held by the working group, the definition being proposed was that 'output is everything to which the agency devotes time and resources'. However, the group soon came to the conclusion that this definition was far too broad and that the conception of output had to take the form of more distinct, commensurable categories. To reduce the uncertainty as to what constitutes an output, the group sought to produce a common definition of this concept that would be applicable across all operating areas within the agency. The quest for such a common definition of output was mainly driven by the Chief Accountant, who was relatively new to the SEA. Unlike most civil servants in the SEA, the Chief Accountant had an administrative career from other parts of the public sector behind him and espoused professional values that were more conducive to the quantification of performance reporting to preserve a sense of administrative order. These values were manifest in a pronounced commensuration mindset. Throughout the process of defining output, the Chief Accountant consistently sought to arrive at a clear-cut, generic understanding of this concept, that would be possible to translate into comparable categories and metrics across various operating areas. On several occasions, he explained how the categorization work undertaken to this end was necessary to overcome the ingrained tendency of viewing the agency's operations as a set of incommensurable singularities (Bialecki et al., 2017; Karpik, 2010) and simplify the reporting. Contrasting this ingrained mindset with his efforts to produce comparable output categories he explained:

R3: People here have learnt that what they are doing is very complex, very different and not simple. We have 84 per cent academics who work on very advanced things and it doesn't really fit within the format [i.e., the emerging set of comparable output categories] [...] First of all, we have very complex tasks, but we also have things that are very *different* [emphasizing the word] from each other. The SEA has a range of different tasks and the reporting has not been streamlined. [...] It's almost as if there were many organizations within the organization. (Interview, 2010-12-02)

We initially found this emerging pursuit of commensurable categories as a basis for performance reporting puzzling. Even though the emerging categorization work can be seen as a response to the need for uncertainty reduction, that was championed by an individual with a pronounced commensuration mindset (cf. Espeland, 1998; Espeland & Stevens, 1998), there was nothing in the Government's evolving reporting requirements that mandated Swedish government agencies to make their performance reporting comparable across operating areas. While the SEA's categorization work emerged as a reactive response to these requirements, it was mainly the Chief Accountant's insistence on the need for comparable categories of outputs that caused the performance reporting to take the form it eventually did. However, as we started to probe the reasons for the nascent categorization work, we discovered that it did not only constitute a direct reaction to the Government's reporting requirements, but that it could also be traced to the evaluation procedures administered by the NAO and the SNFMA that exercised an important, indirect influence on this work.

Insofar as the NAO was concerned, the SEA initially received little guidance as to what might constitute an acceptable reporting format and this reinforced the sense of uncertainty as to how to structure its annual reports. While the NAO had previously been critical of the agency's extensive use of narrative reporting, it was accustomed to follow the activity structure in the appropriation directives as the basis for its audits. In the absence of this structure, it was not in a position to offer firm guidelines as to how the annual reports should be constructed. Following the split between the NAO and the SNFMA in 1998, the auditors were also reluctant to assume an advisory role and offer normative guidelines for how Swedish government agencies should structure their performance reporting. Reflecting on how the resultant uncertainty affected the nascent categorization work in the SEA, the Chief Accountant explained:

R3: I hope the auditors understand that this is not easy the first year [...] When I talked to the auditors, they said that 'this thing [about output] is not a transitory rule' so perhaps we'll get a remark. We don't feel safe since we have no other system [for performance reporting]. I don't feel we are quite there yet where I can say 'that's it, I can really vouch for these numbers'. (Interview, 2009-10-01).

The above quote shows how the uncertainty related to the NAO's auditing practices reinforced the concerns about getting the reporting format right which, in the mind of the Chief Accountant, was more or less synonymous with finding a common definition of output that could be translated into comparable categories. Similar concerns, reinforcing the SEA's categorization work, were notable in relation to the Financial Management Evaluations administered by the SNFMA and the legitimacy that these evaluations bestow on Swedish government agencies. According to the Chief Accountant, the SEA 'had always had a top position' in the ratings embedded in these evaluations and, as the person in charge of the agency's annual report, he was afraid of being downgraded if the agency somehow got the reporting format 'wrong' and this resulted in remarks from the NAO. The importance of maintaining such a position is underscored by the construction of these ratings. The ratings are aggregated into an overriding, nine-point scale, ranging from AA to CC, where the majority of the agencies (usually over 60 per cent) generally receive an AA score while very few are rated lower than BA (the fourth step down from the top). Hence, anything but a 'top position' in these ratings signals a deviation from the norm, that often triggers demands to rectify problems in underlying administrative routines. Unless such problems are

addressed in a timely manner, the agencies run the risk of being reprimanded and, in the long term, losing legitimacy.

The findings reported above show that the initial categorization work in the SEA, which emerged as direct reaction to the Government's changing reporting requirements, was reinforced by an element of tracing work whereby key organizational actors connected the agency's performance reporting to a broader assemblage of calculative infrastructures that bestow legitimacy on Swedish government agencies. Rather than emerging in an infrastructural vacuum, the categorization work initiated by the Chief Accountant was combined with concerted efforts to justify this work with reference to existing performance evaluation procedures such as the NAO's audits and the Financial Management Evaluations. Through this tracing work, connections were forged between the SEA's reporting practices and a broader assemblage of calculative infrastructures at an early stage and this created an additional, indirect layer of reactivity. This indirect reactivity was reinforced by the inter-related nature of the NAO's audits and the Financial Management Evaluations, which makes it difficult to disentangle the uncertainty related to the former and the legitimacy concerns related to the latter as forces influencing the agency's categorization work. What is clear, however, is that their combined effects added impetus to this work as a basis for getting the performance reporting right. As we shall see later in the empirical analysis, the salience of this indirect layer of reactivity increased as organizational actors engaged in further tracing work and, especially, as the SEA came under increasing pressure to legitimize its accounting practices to the NAO.

#### *Continued Commensuration Work and Contestation (Late 2009 to Early 2010)*

While the SEA's initial categorization work was buttressed by calculative infrastructures, such as the NAO audits and the Financial Management Evaluations, these infrastructures did not provide sufficient guidance as to what might constitute more specific output categories and metrics. In its quest for such guidance, the agency engaged in additional tracing work whereby connections to other infrastructures were being forged. The working group first turned to guidelines issued by the SNFMA for how to interpret the intentions behind the Performance Management Investigation which read as follows:

Output refers to products or services generated by the agency's activity. Output refers to an activity that is settled and not a work in progress. The products or services have to have left the agency (e.g., reports and decisions) or be open and available to the general public (e.g., exhibitions and web pages). (Swedish National Financial Management Authority, 2009, p. 15).

The SNFMA's guidelines provided a point of reference to which the definition of outputs could be traced and also a terminology that might help agencies to track outputs back to underlying activities. However, as a basis for identifying more specific, commensurable output categories, the guidelines were initially considered too vague. Hence, as 2009 was drawing to a close and the SEA had to determine the format of its annual report, the working group realized that it would not be able to reduce the uncertainty about how to produce comparable output categories and metrics unless more concrete guidance was forthcoming.

To address these issues, the SEA invited the SNFMA to hold a workshop for agency staff in late 2009 which effectively led to an intensification of its tracing work. During the workshop, the SNFMA re-iterated its definition of output as products or services that 'have left the agency' but the civil servants within the SEA now had the opportunity to discuss the meaning of this definition in greater detail and deepen their understanding of how specific output categories could be traced to underlying activities. In the workshop, the SNFMA went to considerable lengths to explain how outputs are distinct from but related to inputs, activities and outcomes using a model that is well-established in the public sector accounting literature. It also provided

examples of specific output categories to further clarify what this notion could be and explained that output could be ‘services or products the agency provides to achieve its goals that “leave the building,” such as a report, a prescription or a paid grant’.<sup>2</sup> In translating this definition of output into its ongoing categorization work, the working group within the SEA followed the SNFMA’s formulation relatively closely. The notion that something had to have ‘left the building’ (ie. the agency) to count as output was seen as a breakthrough because it offered a common definition that would capture a broad range of operations and be possible to translate into categories that would be comparable across most of the agency’s departments.

The discussion above shows how the SEA’s tracing work, that connected the agency’s emerging reporting practices to the SNFMA’s guidelines for the definition of outputs, buttressed its categorization work. However, the definition of output in terms of things that have ‘left the building’ soon created new problems that exacerbated the commensuration work. Once this common definition of output had been agreed upon, more detailed categorization work aimed at linking specific outputs to the agency’s everyday operations had to be undertaken. Each department within the agency was given the task of providing a list of output-generating operations to the working group responsible for defining outputs. To the surprise of administrative staff, this exercise resulted in a list of over 300 types of operations being proposed as potentially relevant outputs. Consistent with the tendency among agency staff to view their operations as incommensurable singularities, many of these outputs were highly specific to particular operating areas. These problems were exacerbated by the difficulties of tracing many of the outputs to the agency’s time-recording system and, thereby, complying with the requirement that all outputs should be linked to costs. Since the agency was primarily using the time spent on different activities as a basis for cost allocations, the codes for recording the civil servants’ time had to be relatively general in character and could not be related to too many output categories. Hence, the commensuration work in the SEA reached a new impasse. The Chief Accountant recalled:

R3: The list was huge; it was an enormous amount of outputs, and our Director General said: ‘We can’t have this; we can’t present all of this in the annual report’. (Interview, 2010-03-08).

To resolve this problem and translate the agency’s categorization work into metricization work, aimed at producing a manageable number of output indicators, further tracing work connecting outputs to the agency’s time-recording system was required. Mobilizing this system as a supporting, calculative infrastructure, the Director General and the Chief Accountant finally managed to couple outputs to costs and condense the number of outputs to a list of eight, measurable categories representing objects that ‘left the building’ (see Table 1). The Chief Accountant also engaged directly with each department to impress the importance of employees carefully tracing their time to the eight output categories.

The new output categories and metrics gave the commensuration work in the SEA more tangible form. The majority of the outputs listed in Table 1 were applicable across most or all of the agency’s operating departments. However, the output categories and metrics emerging from the tracing of outputs to the agency’s time-recording system encountered a nearly immediate backlash and caused the commensuration work to be hotly contested. As the civil servants started to use the new codes for time reporting, it turned out that many of them were struggling to allocate their time to any of the eight outputs. This made a relatively small proportion of the agency’s operations visible and a category called ‘no output’ was introduced to capture the residual. The creation of this residual category caused considerable frustration. Many of the civil servants also

---

<sup>2</sup>Extract from the Power-Point presentation used during the workshop.

**Table 1.** The list of outputs included in the SEA's annual report for 2009.

	Output category	Volume metrics	Cost metrics
1	Funding decision; concerns all administration of grants and loans through the point at which a decision is made.		
2	Presentations; external teaching and external conferences: concerns meetings, conferences, seminars, etc., that have been organized by the SEA. The prerequisite is that the SEA has organised the presentations, hosted the external lectures or arranged the external conferences, etc. If the agency just participates in a meeting or conference, that does not count as output.		
3	Supervision; concerns the supervision of a certain energy efficiency program for industries administered by the SEA, the emergency storage of oil and carbon, municipal energy planning, energy-marking and eco-design.		
4	Tests; concerns tests of electronic equipment for households and industry.		
5	Decisions regarding electricity certificates; concerns formal decisions regarding electricity certificates as well as examinations of electric declarations.		
6	Completed Clean Development Mechanism (CDM) agreements; concerns agreements regarding climate projects in developing countries.		
7	Publications; concerns published publications from the SEA.		
8	Material provided to other central agencies and international organizations; concerns registered documents sent to the government, government offices, other central agencies or international organizations.		
	<i>Sum</i>		

reacted negatively to what they saw as a rather pejorative, if not insulting, labeling of the 'no output' category.<sup>3</sup> The Chief Accountant recalled:

R3: It was a big mistake calling the last category 'no output'; it created a lot of frustration. People were asking 'do you not think I am doing anything'; so now we call that category 'other activity' instead. (Interview, 2010-03-08).

However, the new label did not help much. When the list of outputs was finally published in the annual report for 2009, approximately 40 per cent of the civil servants' total working hours were included under the eight output categories while the remainder fell under the 'other activity' category. The two main areas that were excluded from the eight output categories were operations known internally as 'facilitating operations', which reflect the agency's work in supporting industry and local and regional government in developing efficient and sustainable energy systems, and 'support to the government'. The exclusion of especially the latter area caused significant unease as it reflects the SEA's key role as an expert agency supporting the development of energy policy and led to an open conflict between actors espousing diverging values. In contrast to the Chief Accountant's efforts to simplify performance reporting by increasing the reliance on comparable output metrics, civil servants, who considered themselves experts in this area, wanted to include as much information as possible about their operations and started to question the commensuration work. In doing so, they drew attention to alternative performance aspects that were not captured by the eight output categories. One of our interviewees, who spent a significant part of his working time on 'support to the government', said:

<sup>3</sup>The formulation in Swedish was 'ingen prestation' which can also be translated as 'making no effort' or 'no achievement'.

R8: If some of our activity is not included, there is a risk [...] the annual report shall also send a signal back to us, about activity, how efficient it has been for example [...]. And the greatest risk of all is that an activity that is not visible might over time cease to exist, and somewhere in the future, money comes into the picture. As an example, we worked very hard with the climate negotiations [in conjunction with the 2009 United Nations Climate Change Conference in Copenhagen]. We had seven to eight people working with this last year, and each of them had 200 h overtime or more. And there is not a single trace of that here [in the list of outputs]. [...] I think it's stupid it is not included in the output categories. Nothing has been such a politically hot topic as the climate issue. If you ask the Ministry of the Environment, we get a lot of credit for this. They even wrote us a letter of appreciation for our work in Copenhagen. (Interview, 2010-03-12).

Similar sentiments, showing how the failure to account for a large proportion of the SEA's operations as commensurable outputs exacerbated conflicts, were expressed in one of our other interviews with a civil servant who mainly worked with 'support to the government':

R11: This has been a painful process [...] I would almost like to say that it has been fraught with conflict. There have been many upset feelings [...]. When we eventually understood what an output was, we realized that the majority of the activity in my department does not fit any output type. That is not good. Of course, you want your activity to be accounted for in the annual report [...] now it becomes 'other activity' [...]. I have to say that we have had a very frustrated atmosphere. I think it is extremely sad that we do not account for this, especially our supporting operations to the government. It does not show [in the output categories] how much time we spend on this. [...] we are supposed to always be available when the Government needs us, but we can't account for the time we spend on this. And then, when we shall ask for more resources [...] there is a risk that external actors get the wrong picture of our activity, they might ask 'how do they spend the rest of their time?' (Interview, 2010-03-31).

The two quotes above show that the objections to the exclusion of performance aspects with high policy relevance were not only rooted in the deeply held professional values of civil servants in the SEA, but also in concerns with material aspects such as the potential loss of funding due to the failure to account for such aspects. However, especially the first of these quotes testifies to a strong sense among the civil servants that, in their capacity as experts, they were in a superior position to determine what information the Government needed to further the development of Swedish energy policy. The ability to provide such information was reportedly seen as key to their sense of living up to their professional values:

R9: It is a very important aspect [...] one of the most important reasons for working here is to be a part of changing the energy system, it is a strong driving force, we see that in our employee satisfaction surveys. (Interview, 2010-03-24).

The view that the SEA's commensuration work risked suppressing the provision of policy-relevant information was corroborated in our interviews with civil servants from the MEEC. One of them explained:

R17: What we are interested in is to be able to account [to Parliament] in relation to the objectives for energy policy. [...] We are not interested in how many reports that 'leave the building' and we are not interested in how many decisions are made or how long the average decision takes because this is a kind of operation where one report may be worth just as much as 30 other [reports] and one decision can be worth just as much as 40 other decisions. This kind of quantitative reporting is quite irrelevant for us. [...] For a facilitating agency like the SEA, we are more interested in the effects [of policy objectives] in the longer term. [...] The Government is accountable to Parliament for the objectives for energy policy and the SEA is one of our most important vehicles in this regard. (Interview, 2010-03-19).

The above quote shows how commensurable metrics, such as the number of decisions and publications, not only detract from other performance aspects that are more relevant from the point of view of policy development, but also conceal unique, incommensurable properties such as the perceived worth of individual decisions or publications. Such concerns with incommensurables also featured in our interviews within the SEA and suggest that the discontent with the agency's commensuration work was not confined to civil servants whose work was excluded from the eight output categories. Other civil servants, whose work *was* covered by these categories,



drew attention to how comparisons based on this categorization might conceal incommensurable properties of specific outputs and how this also became a source of contestation. One of them explained:

R15: You can work with one publication for a whole year, and one publication can cost several millions; there can be a lot of actors involved and you have had to consider a huge amount of various issues, but still it is only one [output] [...] it will be weird to examine the number of publications and their costs, it has to be very clear [in the annual report] that the output categories do not provide the whole picture [otherwise] someone might think 'oh, this was an expensive publication'. (Interview, 2010-04-29).

The emerging contestation of the commensuration work prompted administrative staff to take further initiatives to deal with resistance and resolve the conflicts around the agency's performance reporting. As explicated below, this entailed a significant amount of tracing work whereby administrative staff not only continued to depend on the SNFMA's guidelines for the definition of outputs to legitimize their position but also other infrastructures, such as the NAO audits, to develop the agency's performance reporting.

### *Continued Contestation and Resolution (Early 2010 to Late 2011)*

While the tracing work connecting the SEA's reporting practices to the SNFMA's guidelines for the definition of outputs was initially confined to the establishment of an external reference point for the agency's categorization work, it was increasingly turned inwards and used as a political resource in the efforts to overcome resistance. For instance, in one of our earlier interviews the Chief Accountant exclaimed:

R3: The whole agency has been discussing this, 'how the hell shall we do this' [i.e., arriving at a common definition of output] [...] staff feel like this is a burden laid on them from the accounting unit, but we did not make this up. It is the SNFMA, not us; it is a requirement. I say that 'this has to be included, it is a regulation, and it is not a free choice'. (Interview, 2010-03-08).

Interestingly, the Chief Accountant is here forging an explicit connection between the SEA's reporting practices and the SNFMA's guidelines while construing compliance with these guidelines as a mandatory requirement, ostensibly in an attempt to legitimize his position vis-à-vis agency staff. Similar tracing work aimed at legitimizing the position of administrative staff was repeatedly mobilized as the working group in charge of developing the agency's performance reporting continued its work. In early 2010, the working group held four meetings aimed at extending its categorization work to 'facilitating operations' and 'support to the government'. An extract from the discussion at one of the meetings is reproduced below and shows how the tensions between administrative (R4) and operations-focused staff (R20, R21, and R22) played out. While the former persisted in tracing the format of the agency's performance reporting to the SNFMA's guidelines and construing these guidelines as mandatory, the latter challenged this position and continued to draw attention to the need to make alternative and not necessarily commensurable performance aspects visible:

R20: We work a lot with the assignments that lead to a report. The report will be registered and journalized. But [what about] the support to the government that is not registered and journalized? It could be that we negotiate on behalf of the Government or write a memo to the government official who carries out the negotiations. In my department, we currently discuss starting to register and journalize this, to include it [in the output categories] [...] but this is a lot of work.

R4: The reason why we demand registering and journalizing is because we have to know when an output occurs. We have to narrow it down as much as possible. If we widen the definition, we lose our sense of what an output is.

R21: If so much of what we do is not output, then why should we account for output if it does not say anything about what we do?

R4: It is decided that all of the Government shall be governed in this way, and we just have to accept it [referring to the SNFMA's guidelines].

R21: We can't just accept it. We have to find a way that works for our operations, and we cannot just exclude a large part of our operations just because the SNFMA has decided that we shall construct our performance reporting in this way.

R4: No...

R21: I am only saying that we have to adjust this [the output definition]. We cannot get stuck with something that leads to 60 percent of our activity disappearing.

R22: We are supposed to provide a fair view of our activity.

R20: Exactly, we decide what a fair view is.

R21: We can't just register and journalize everything we do just to be able to account for output. (Meeting observations, 2010-04-28).

Because of the difficulties of expanding the definition of output, the working group discussed the possibility of using other concepts such as 'activities' as a complement to the eight output categories. However, as indicated by the meeting observations below, this suggestion was also rejected by the Chief Accountant (R3) with reference to the need to comply with the SNFMA's guidelines. Interestingly, these observations also show how at least some operations-focused staff (R20) began to engage in similar tracing work by invoking these guidelines as a reference point for what could be included in the annual report:

R7: Maybe we should have a category [in the annual report] that we call 'activities' to provide a fair view of our activity?

R3: The SNFMA does not allow for the use of activities in the annual report.

[...]

R20: The SNFMA is very clear about what an activity is. An activity is something that leads to output. Maybe we should use another word? (Meeting observations, 2010-05-05).

Similarly, where alternative terminologies to that of 'activities' were being considered as a basis for expanding the performance reporting, this was also framed in terms of what would be permissible in relation to the SNFMA's guidelines by administrative as well as operations-focused staff:

R20: Is it possible to use 'task', or does 'task' also need to lead to output [in the SNFMA's guidelines]?

R4: We have 'tasks' in our instruction.

R20: I just wonder if 'tasks' are included in this chain [the depiction of the transformation of activities into outputs in the SNFMA's guidelines], do 'tasks' have to lead to output? (Meeting observations, 2010-05-05).

While it would be going too far to suggest that operations-focused staff began to adopt a commensuration mindset, similar to that displayed by administrative staff, our meeting observations show that the SNFMA's guidelines became an increasingly shared reference point for the meeting participants and how this cemented the connections between these guidelines and the agency's reporting practices. This increasingly pervasive tracing work occurred without the SNFMA changing its non-prescriptive stance concerning how Swedish government agencies should structure their performance reporting. However, it is perhaps worth noting that the recurring recourse to such tracing work emerged at a time when these guidelines began to gain traction as a new, voluntary standard for their performance reporting (Svärdsten, 2015). This may have contributed to reducing the perceived uncertainty surrounding the SEA's performance reporting and convinced staff that they were acting in a legitimate manner. Throughout our interactions with especially administrative staff, it became clear that the SNFMA's guidelines were not just seen as a useful means of legitimizing the commensuration work within the agency, but also for

enhancing their sense of being compliant with external accounting regulations and thus avoiding the risk of losing external legitimacy. Reflecting on the strongly felt need for such compliance and how it was facilitated by the commensuration work initiated within the SEA, the Chief Accountant explained:

R3: When you do something like this [i.e., the new performance reporting] and want to be able to stand in front of the auditors [i.e., the NAO] and say that it is done correctly, it is hard to account for all the soft values because you have to be able to tell the auditors that 'yes, everyone is reporting the same way' [...] You can't account for everything whichever way you like. You've got to have fairly sharp boundaries. [...] If you are being audited and people report in different ways, there will be problems. (Interview, 2010-12-02).

Similar to the findings reported in the first phase of our analysis, the above quote shows how the mobilization of tracing work to connect the evolving reporting practices to the calculative infrastructures provided by the NAO and make these connections visible reinforced the indirect layer of reactivity influencing the commensuration work in the SEA. By connecting the emerging reporting practices, emanating from the use of the SNFMA's guidelines, to these infrastructures this tracing work offered a degree of reassurance that the agency's performance reporting was appropriate. Similar sentiments, showing how consequential, but also challenging, this tracing work ultimately was were expressed by another interviewee:

Interviewer: Has the extent of documentation increased [due to the commensurations work]?

R5: Well, yes, since we need to have the documentation, such as the presentations for example, that the NAO wants. [...] The NAO wants every presentation to be saved. I think there was one department that had saved the documents and another one that tried to pull everything together, but it was impossible. [...] We were panicking a bit to pull everything we had presented together for the annual audit. In the end, we decided that it was important to, for example, have the program for the conference to show that we have actually been there if we can't get hold of the presentations. [...] The demands for documentation have become completely different. (Interview, 2011-12-15).

However, the same interviewee also lamented how the strongly felt need to trace the reporting back to what was expected by the NAO reinforced the tendency for the Government's information needs to be displaced:

R5: What we really would like to account for is what we do to help the Government but that is not registered.

Interviewer: Have you not started to document this?

R5: No, and this is the problem all the time. How do we account for volume? [...] It comes back to volume [of outputs] all the time. (Interview, 2011-12-15).

During the spring of 2010, the influence of the NAO became more acutely felt as the auditors raised a number of critical remarks on the SEA's annual report for 2009. Even though the remarks were not related to the problems associated with the agency's reporting of outputs<sup>4</sup>, they were followed by a drop in its rating from AA to BA in the Financial Management Evaluation. These events enhanced the sense of urgency about getting the performance reporting right which was palpable in our interviews and interactions with administrative staff. For instance, during a field visit shortly after the criticisms by the NAO were made public, the first author met a very stressed Chief Accountant who was on his way to the Director General to discuss these criticisms. In an interview immediately after this encounter another member of staff underlined the severity of the matter and the strongly felt need to be compliant with external accounting regulations:

R9 They [the National Audit Office] were fairly satisfied [with the reporting of outputs] given the circumstances and the short time we had available [...] but then we got some remarks about other things.

Interviewer: He [the Chief Accountant] just mentioned that.

---

<sup>4</sup>The remarks mainly concerned the treatment of current assets in the agency's balance sheet.

R9: Yes, that definitely wasn't good! (Interview, 2010-03-24).

Insofar as the commensuration work in the SEA was concerned, the growing pressures to legitimize the agency's accounting practices to external parties that followed from these events reinforced the reluctance of administrative staff to change the reporting format emerging from their earlier work. Given that the NAO had approved of the format of the annual report for 2009, the Chief Accountant did not wish to open up the discussion about how output should be defined again and continued to insist that the reporting of output metrics should follow the same format to avoid the risk of getting additional, critical remarks from the auditors and another unfavorable rating in the Financial Management Evaluations. These observations show how the efforts to connect the SEAs evolving performance reporting to a broader assemblage of calculative infrastructures were reinforced by the legitimacy threat emerging in the spring of 2010. This tracing work effectively buttressed the indirect layer of reactivity influencing the categorization work in the SEA as it contributed to the decision to maintain the existing categorization of outputs. However, this course of action, which effectively constrained the agency's performance reporting, only fueled additional conflicts. Throughout its meetings during the spring of 2010, the working group in charge of developing the agency's performance reporting failed to arrive at a consensus solution. It was only as the year was drawing to a close that a compromise between the use of comparable output metrics and the narrative reporting format that had historically dominated the annual reports began to emerge. Describing how he had somewhat reluctantly agreed to this compromise, the Chief Accountant explained:

R3: I would say that we have had two alternatives [for resolving the conflicts emerging from the commensuration work]. One is to relax the boundaries for what counts as an output according the SNFMA's guidelines and we haven't thought this is a great idea. The other alternative, which we will pursue, is to describe all the other operations [that are not included in the eight output categories] in the form of text. In other words, we are trying to incorporate all this through the back door [sighs and catches his breath] through [Microsoft] Word because it didn't feel right to relax the boundaries [emerging from the commensuration work]. (Interview, 2010-12-02).

To some extent, the reporting format that evolved from the compromise relaxed the emphasis on commensuration. In contrast to the annual report for 2009, the report for 2010 did not provide a comparison of the output metrics across different departments on a stand-alone and largely de-contextualized basis. Instead, the metrics were dispersed across different chapters of the report where they were accompanied by extensive narratives describing the agency's operations. The use of narratives was seen as alleviating the risks of not being able to provide policy-relevant information and external audiences drawing the 'wrong' conclusions about the agency's operations. Also, the annual report for 2010 clearly stated that the eight output categories emerging from the agency's categorization work did not reflect all of its operations and that other performance aspects might also be relevant to take into consideration. However, the agency could still account for outputs in a way that was seen as following the SNFMA's guidelines. The output metrics emerging from the agency's metricization and tracing work also continued to be tightly coupled to people's time reporting and the allocation of costs.

The compromise outlined above eventually caused the conflicts around the SEA's performance reporting to abate and led the reporting practices to achieve a new state of stability. The agency's annual report for 2011 followed a similar format as the one for 2010 and continued to gain the approval of the NAO. As such, the unfolding categorization and tracing work enabled the SEA to avoid further legitimacy threats, such as those emerging in the spring of 2010, and eventually satisfied a broad range of interests and values within and beyond the agency. Ultimately, the commensuration work unfolding in the SEA thus preserved the heterarchy of values (Lamont, 2012) that became salient in this work. However, it is questionable whether the SEA took full advantage of the discretion that the Government sought to bestow on Swedish government agencies. In our informal conversations with two of the civil servants in charge of the Performance

Management Investigation, which extended throughout our field work, they expressed severe frustration about the way the agency approached the reporting of outputs. In their view, Swedish government agencies were now free to devise their performance reporting whichever way they liked and they struggled to understand why the agency deliberately constrained these possibilities by attaching so much weight to external guidelines and regulations. Yet, the SEA was not the only agency that struggled with the reporting of outputs. Indeed, in a broadly based investigation conducted in 2011, the SNFMA concluded that in most of the agencies being consulted ‘the requirement to account for output (according to the standard definition)<sup>5</sup> was questioned and regarded as constraining, irrelevant or even impossible to carry out’ (Swedish National Financial Management Authority, 2012, p. 39). The SNFMA also concluded that several ministries did no longer receive policy-relevant performance information. Hence, the application of the SNFMA’s standard definition of output, which buttressed the commensuration work in the SEA, seems to have created similar problems to the ones we have observed in many other agencies.

### **Concluding Discussion**

In this paper, we have asked how organizations’ commensuration work unfolds where a broader assemblage of calculative infrastructures underpins societal reform programs. We have addressed this research question in the context of Swedish central government during a period when a societal reform program aimed at enhancing the autonomy of government agencies while increasing their accountability for results was amplified by an attempt to radically increase their reporting discretion. In doing so, we have nuanced the view that the pursuit of commensuration by evaluated organizations manifests itself as a reactive response to a single, externally sanctioned set of clearly identifiable and comparable metrics which has dominated prior research on the topic (Pollock et al., 2018; Slager & Gond, 2022). Instead, what is emerging from our empirical analysis is a multi-layered view of commensuration work that recognizes that direct, reactive responses can interact with more indirect forms of reactivity that can be traced back to a broader assemblage of calculative infrastructures. In the SEA, the inter-related infrastructures provided by the NAO and the Financial Management Evaluations clearly reinforced the commensuration work unfolding within the agency and contributed to stabilizing the emerging performance reporting practices even though they were not the direct progenitors of these practices. Hence, calculative infrastructures can buttress organizations’ commensuration work and reinforce tendencies toward reactivity even though they are somewhat removed from the accounting practices that are subject to such work. As we have also seen, this entails a reciprocal relationship between commensuration and reactivity where reactive responses do not follow from a set of readily available, commensurable metrics but rather evolve as such metrics get connected to a broader assemblage of calculative infrastructures and lead to further commensuration work.

We extend this reciprocal, multi-layered view of commensuration work by connecting it to three distinct, but inter-related types of work, namely, categorization work, metricization work and, especially, tracing work as a basis for explaining how such work unfolds in relation to a broader assemblage of calculative infrastructures. In the SEA, the categorization and metricization work, that emerged as a direct, reactive response to the Government’s changing reporting requirements, was supported by tracing work whereby the emerging performance reporting practices were connected to other calculative infrastructures such as the SNFMA’s guidelines for the definition of output and the agency’s time-recording system. The agency’s tracing work played

---

<sup>5</sup>This standard definition refers to the one found in the SNFMA’s guidelines for performance reporting that the SEA used as a basis for defining outputs and producing comparable output metrics.

an equally important role in connecting these reporting practices to the infrastructures provided by the NAO and the Financial Management Evaluations and thus mediated the indirect layer of reactivity that was evident in its commensuration work. The significance of this tracing work is perhaps best understood in the light of the uncertainty that surrounded the Government's reporting requirements following its ambition to enhance the reporting discretion of Swedish government agencies. Given the Government's reluctance to issue detailed regulations prescribing how the agencies should structure their performance reporting, it was difficult for the SEA to trace its reporting needs back to these requirements. Instead, to reduce uncertainty and legitimize its emerging reporting practices within as well as beyond the agency, it turned to calculative infrastructures provided by other actors that bestow legitimacy on Swedish government agencies. These infrastructures subsequently buttressed the emerging reporting practices and offered an important impetus for the agency's commensuration work.

These findings support Power's (2019) observation that the ability of organizations to connect their accounting practices to a broader assemblage of calculative infrastructures can play an important role in the stabilization of such practices. However, as we have also seen, tracing work that aims at forging such connections can reinforce the risks of organizations displacing the focus on their overriding missions and fuel contestation. In the SEA, this risk was especially evident as its performance reporting practices not only constrained the reporting discretion that the Government sought to bestow on Swedish government agencies, but also jeopardized the provision of policy-relevant information. This, in turn, caused the agency's commensuration work to be hotly contested. These problems only emerged as the agency began to engage in tracing work to support its categorization and metricization work and were subsequently exacerbated as the agency increasingly traced its performance reporting practices to the demands of the NAO rather than the Government's information needs. These findings echo Power's (2021) observation that the growing demands on organizations to trace their accounting practices to a broader assemblage of calculative infrastructures can easily lead to various forms of goal displacement unless such infrastructures are aligned with their principal objectives. Tracing work should thus be seen as a potentially double-edged sword which, on the one hand, can stabilize organizational accounting practices by reinforcing tendencies toward reactivity while, on the other, fueling contestation that disrupts such practices. In the SEA, a compromise that allowed the agency to strike a balance between these opposing forces was only reached after a protracted period of commensuration work that ultimately preserved its heterarchy of values.

The findings reported above offer two distinct insights that underline the importance of linking the notion of commensuration work to the literature on calculative infrastructures. First, in contrast to prior research on commensuration that has started to pay attention to the impact of multiple performance evaluation procedures, such as competing rankings or ratings (Dorn, 2019; Pollock et al., 2018; Sauder & Espeland, 2006; Slager & Gond, 2022), we show that the existence of a broader assemblage of infrastructures can reinforce rather than weaken the tendencies toward reactivity even though such tendencies are contested. Whereas prior research has shown that the propensity of organizations to comply with a single set of comparable metrics is reduced when they can choose between diverse rankings or ratings, we demonstrate how calculative infrastructures buttress reactive tendencies but also that this reactivity forms more complex patterns than what has been documented in the literature on commensuration. As our findings from the SEA make plain, this complexity can only be accounted for by widening the conception of calculative infrastructures and understanding how organizations engage in tracing work that connects their accounting practices to such infrastructures. This insight is more consonant with the emerging literature on calculative infrastructures (Habersam et al., 2021; Juven, 2019; Kornberger et al., 2017; Kurunmäki et al., 2019; Kurunmäki & Miller, 2013; Power, 2015; Reilley & Scheytt, 2019) and, especially, Power's (2019, 2021) closely related work on traceability, than prior research on

commensuration and reactivity due to the latter's focus on a single set of comparable metrics or, at best, a limited range of competing performance evaluation procedures.

Second, by linking the notion of commensuration work to the literature on calculative infrastructures we offer novel insights into how unintended consequences, that run counter to the ambitions behind societal reform programs, emerge. The literature on commensuration is replete with evidence of unintended consequences of performance evaluation (Espeland & Sauder, 2007, 2016; Kornberger & Carter, 2010; Llewellyn & Northcott, 2005; Sauder et al., 2021). However, due to its neglect of how commensuration work unfolds in relation to a broader assemblage of calculative infrastructures, it often fails to account for the wider, systemic effects that can emerge from such work. Since calculative infrastructures are often of a distributed nature, in the sense that they are not tightly coordinated by a single actor (Kornberger et al., 2017; Reilley & Scheytt, 2019), it is quite likely that commensuration work will have less predictable outcomes that jeopardize the ambitions behind societal reform programs compared to situations where organizations only react to a limited set of metrics that is sanctioned by the champions of such programs. As our findings from the SEA show, the prevalence of such infrastructures can compel organizations to engage in commensuration work that effectively counteracts the intentions behind societal reform programs and reinforces tendencies toward goal displacement. Following Power (2021), the risk of such goal displacement is greater the more misaligned calculative infrastructures are with the principal objectives of organizations. Also, the likelihood of such misalignment is possibly greater the less coordinated and more dispersed such infrastructures are in particular societal sectors. This reinforces our conjecture that the risk of unintended consequences of commensuration work that counteract societal reform programs is higher where such work unfolds in relation to a broader assemblage of calculative infrastructures compared to situations where this is not the case.

Our findings have important implications for future research on commensuration and the broader literature on performance evaluation in which it is situated. Given the weight we have attached to especially tracing work as a mechanism that supports other types of commensuration work, such as categorization and metricization work, and that mediates indirect forms of reactivity, our study responds to Power's (2019, p. 127) call for developing a program of 'traceability studies'. As our analysis indicates, advancing such a program may be particularly relevant where tendencies toward reactivity are not confined to direct responses to externally sanctioned reporting requirements, but also entail indirect forms of reactivity emanating from a broader assemblage of calculative infrastructures. However, our multi-layered view of commensuration work could be extended to contexts where direct forms of reactivity, originating in clearly identifiable sets of comparable metrics that are forcefully promoted by external actors, are more salient. Relevant questions to explore in such studies include whether direct forms of reactivity will outweigh indirect forms of reactivity, what roles different types of commensuration work play in producing different types of reactivity and whether such work enhances or reduces the risk of unintended consequences. While prior research on commensuration suggests that the external imposition of comparable metrics will reinforce direct forms of reactivity (Espeland & Sauder, 2007, 2016), more research is required into whether such forms of reactivity will continue to dominate when organizations trace their accounting practices to a broader assemblage of calculative infrastructures or whether such tracing work will produce more complex patterns of reactivity similar to those observed in our study. This underlines the need to adopt a comparative perspective that allows researchers to theorize how different forms of reactivity manifest themselves under different infrastructural circumstances. In addition to advancing a broader program of traceability studies (Power, 2019, 2021), such research can respond to emerging calls for further studies of how variations in reactivity emerge and give rise to unintended consequences in the literature on commensuration (Sauder et al., 2021; Slager & Gond, 2022).

A comparative approach to theory development is also warranted to extend insights into how commensuration work unfolds that are perhaps somewhat specific to our research setting. The commensuration work examined in our study evolved in a highly devolved governance context where notions of agency autonomy constitute a long-standing, institutionalized ideal and where the Government sought to enhance this autonomy by increasing the agencies' reporting discretion. As we have seen, this reform agenda created considerable uncertainty and was, in many ways, the root cause for why the SEA became increasingly dependent on a broader assemblage of calculative infrastructures as a basis for its commensuration work. It would be potentially interesting to contrast these observations with how commensuration work unfolds in more centralized governance contexts where reforms evolve in a more pronounced top-down manner. It is possible that such governance contexts are characterized by somewhat tighter coordination of the calculative infrastructures that govern accounting practices and that this gives rise to forms of reactivity and contestation that differ from those observed in our study. However, public sector governance reforms often evolve in serendipitous manners where accounting plays an unpredictable, constitutive role that can reinforce unintended consequences (Dahler-Larsen, 2014; Modell, 2022). Hence, it would be fallacious to assume that commensuration work will inevitably unfold in more orderly and predictable ways in highly centralized governance contexts.

A final comparative aspect that requires greater attention in future research on commensuration work is the degree of professionalization of the organizations where such work takes place and the value conflicts that often follow from such professionalization. In this regard, our research setting represents a more typical case of a highly professionalized organization, where commensuration work greatly amplified underlying value conflicts (cf. Elsbach & Kramer, 1996; Englund & Gerdin, 2020; Espeland, 1998; Gerdin & Englund, 2019; Sauder & Espeland, 2009), although the contestation of this work could also be traced to material concerns such as the risk of reduced funding. Further, comparative research is required to enhance our understanding of how commensuration work differs between such contexts and organizational settings characterized by lower levels of professionalization. In less highly professionalized organizations, where value conflicts are perhaps less salient, commensuration work is possibly less contested and reactive tendencies therefore stronger. However, we still know little about how such tendencies toward contestation and reactivity are affected by a multi-layered notion of commensuration work unfolding in relation to a broader assemblage of calculative infrastructures.

## **Acknowledgements**

Earlier versions of this paper were presented at research seminars at Alliance Manchester Business School, Stockholm University Business School, HEC Paris and WU Vienna and at the 44th Annual European Accounting Association (EAA) Congress, Bergen. We are particularly grateful to Matteo Ronzani, who read and commented on several earlier versions of this paper, and two anonymous *EAR* reviewers. We also acknowledge the helpful comments by Charlotta Bay, Christos Begkos, Thomas Carrington, Bino Catusus, Gustav Johed, Daniel Martinez, Jan Mouritsen, Brendan O'Dwyer, Dane Pflueger, Keith Robson, Bob Scapens, Ileana Steccolini and Eva Wittbom.

## **Disclosure statement**

No potential conflict of interest was reported by the author(s).

## **Supplemental Data and Research Materials**

Supplemental data for this article can be accessed on the Taylor & Francis website, <https://doi.org/10.1080/09638180.2023.2166091>.



## References

- Berg, B. L., & Lune, H. (2012). *Qualitative research methods for the social sciences* (8th ed.). Pearson.
- Bialecki, M., O'Leary, S., & Smith, D. (2017). Judgement devices and the evaluation of singularities: The use of performance ratings and narrative information to guide film viewer choice. *Management Accounting Research*, 35, 56–65. <https://doi.org/10.1016/j.mar.2016.01.005>
- Crvelin, D., & Löhlein, L. (2022). Commensuration by form: Lists and accounting in collective action networks. *Accounting, Organizations and Society*, 100, 101333. <https://doi.org/10.1016/j.aos.2021.101333>
- Dahler-Larsen, P. (2014). Constitutive effects of performance indicators: Getting beyond unintended consequences. *Public Management Review*, 16(7), 969–986. <https://doi.org/10.1080/14719037.2013.770058>
- Dorn, C. (2019). When reactivity fails: The limited effects of hospital rankings. *Social Science Information*, 58(2), 327–353. <https://doi.org/10.1177/0539018419848994>
- Elsbach, K. D., & Kramer, R. M. (1996). Members' responses to organizational identity threats: Encountering and countering the business week rankings. *Administrative Science Quarterly*, 41(3), 442–476. <https://doi.org/10.2307/2393938>
- Englund, H., & Gerdin, J. (2020). Contesting conformity: How and why academics may oppose the conforming influences of intra-organizational performance evaluations. *Accounting, Auditing & Accountability Journal*, 33(5), 913–938. <https://doi.org/10.1108/AAAJ-03-2019-3932>
- Espeland, W. N. (1998). *The struggle for water. Politics, rationality, and identity in the American southwest*. University of Chicago Press.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1–40. <https://doi.org/10.1086/517897>
- Espeland, W. N., & Sauder, M. (2016). *Engines of anxiety. Academic rankings, reputation, and accountability*. Russell Sage.
- Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual Review of Sociology*, 24(1), 313–343. <https://doi.org/10.1146/annurev.soc.24.1.313>
- Espeland, W. N., & Stevens, M. L. (2008). A sociology of quantification. *European Journal of Sociology*, 49(3), 401–436. <https://doi.org/10.1017/S0003975609000150>
- Esposito, E., & Stark, D. (2019). What's observed in a rating? Rankings as orientation in the face of uncertainty. *Theory, Culture & Society*, 36(1), 3–26. <https://doi.org/10.1177/0263276419826276>
- Fourcade, M. (2011). Cents and sensibility: Economic valuation and the nature of “nature”. *American Journal of Sociology*, 116(6), 1721–1777. <https://doi.org/10.1086/659640>
- Gerdin, J., & Englund, H. (2019). Contesting commensuration. *Accounting, Auditing & Accountability Journal*, 32(4), 1098–1116. <https://doi.org/10.1108/AAAJ-09-2016-2717>
- Grönlund, A., Svärdesten, F., & Öhman, P. (2011). Value for money and the rule of law: The (new) performance audit in Sweden. *International Journal of Public Sector Management*, 24(2), 107–121. <https://doi.org/10.1108/09513551111109026>
- Habersam, M., Piber, M., & Skoog, M. (2021). Calculative regimes in the making: Implementation and consequences in the context of Austrian public universities. *Qualitative Research in Accounting & Management*, 18(2), 169–194. <https://doi.org/10.1108/QRAM-01-2019-0021>
- Habran, Y., & Mouritsen, J. (2022). Making intensity of efforts the same: Commensuration work in target-setting practices. *European Accounting Review*, 31(3), 603–627. <https://doi.org/10.1080/09638180.2020.1832901>
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19. <https://doi.org/10.1111/j.1467-9299.1991.tb00779.x>
- Hood, C. (1995). The “new public management” in the 1980s: Variations on a theme. *Accounting, Organizations and Society*, 20(1), 93–109. [https://doi.org/10.1016/0361-3682\(93\)E0001-W](https://doi.org/10.1016/0361-3682(93)E0001-W)
- Huault, I., & Rainelli-Weiss, H. (2011). A market for weather risk? Conflicting metrics, attempts at compromise, and limits to commensuration. *Organization Studies*, 32(10), 1395–1419. <https://doi.org/10.1177/0170840611421251>
- Jeacle, I. (2022). The gendered nature of valuation: Valuing life in the titanic compensation claims process. *Accounting, Organizations and Society*, 99, 101309. <https://doi.org/10.1016/j.aos.2021.101309>
- Jeacle, I., & Carter, C. (2011). In TripAdvisor we trust: Rankings, calculative regimes and abstract systems. *Accounting, Organizations and Society*, 36(3), 293–309. <https://doi.org/10.1016/j.aos.2011.04.002>
- Juven, P.-A. (2019). Research in the sociology of organizations. *Research in the Sociology of Organizations*, 62, 69–84. <https://doi.org/10.1108/S0733-558X20190000062005>
- Karpik, L. (2010). *Valuing the unique: The economics of singularities*. Princeton University Press.
- Kolk, A., Levy, D., & Pinkse, J. (2008). Corporate responses in an emerging climate regime: The institutionalization and commensuration of carbon disclosure. *European Accounting Review*, 17(4), 719–745. <https://doi.org/10.1080/09638180802489121>

- Kornberger, M., & Carter, C. (2010). Manufacturing competition: How accounting practices shape strategy making in cities. *Accounting, Auditing & Accountability Journal*, 23(3), 325–349. <https://doi.org/10.1108/09513571011034325>
- Kornberger, M., Pflueger, D., & Mouritsen, J. (2017). Evaluative infrastructures: Accounting for platform organization. *Accounting, Organizations and Society*, 60, 79–95. <https://doi.org/10.1016/j.aos.2017.05.002>
- Kurunmäki, L., Mennicken, A., & Miller, P. (2019). Assembling calculative infrastructures. *Research in the Sociology of Organizations*, 62, 17–42.
- Kurunmäki, L., & Miller, P. (2013). Calculating failure: The making of a calculative infrastructure for forgiving and forecasting failure. *Business History*, 55(7), 1100–1118. <https://doi.org/10.1080/00076791.2013.838036>
- Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38(1), 201–221. <https://doi.org/10.1146/annurev-soc-070308-120022>
- Llewellyn, S., & Northcott, D. (2005). The average hospital. *Accounting, Organizations and Society*, 30(6), 555–583. <https://doi.org/10.1016/j.aos.2004.05.005>
- Lukka, K., & Modell, S. (2010). Validation in interpretive management accounting research. *Accounting, Organizations and Society*, 35(4), 462–477. <https://doi.org/10.1016/j.aos.2009.10.004>
- Lukka, K., & Modell, S. (2017). Interpretive research in accounting: Past, present and future. In Z. Hoque, L. Parker, M. Covaleski, & K. Haynes (Eds.), *The routledge companion to qualitative accounting research methods* (pp. 36–54). Routledge.
- Mehrpourya, A., & Samiolo, R. (2016). Performance measurement in global governance: Ranking and the politics of variability. *Accounting, Organizations and Society*, 55(1), 12–31. <https://doi.org/10.1016/j.aos.2016.09.001>
- Mennicken, A. (2013). ‘Too Big to fail and Too Big to succeed’: Accounting and privatisation in the prison service of England and Wales. *Financial Accountability & Management*, 29(3), 206–226. <https://doi.org/10.1111/faam.12012>
- Mennicken, A., & Espeland, W. N. (2019). What’s new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology*, 45(1), 223–245. <https://doi.org/10.1146/annurev-soc-073117-041343>
- Miller, P. (1991). Accounting innovation beyond the enterprise: Problematizing investment decisions and programming economic growth in the U.K. In the 1960s. *Accounting, Organizations and Society*, 16(5), 733–762. [https://doi.org/10.1016/0361-3682\(91\)90022-7](https://doi.org/10.1016/0361-3682(91)90022-7)
- Modell, S. (2014). The societal relevance of management accounting: An introduction to the special issue. *Accounting and Business Research*, 44(2), 83–103. <https://doi.org/10.1080/00014788.2014.882741>
- Modell, S. (2021). The (non-)adoption of outcome-based performance management: Evidence from Swedish central government. In Z. Hoque (Ed.), *Public sector reform and performance management in developed countries: Outcomes-based approaches in practice* (pp. 265–285). Routledge.
- Modell, S. (2022). New developments in institutional research on performance measurement and management in the public sector. *Journal of Public Budgeting, Accounting & Financial Management*, 34(3), 353–369. <https://doi.org/10.1108/JPB AFM-04-2021-0070>
- Modell, S., Jacobs, K., & Wiesel, F. (2007). A process (re)turn? Path dependencies, institutions and performance management in Swedish central government. *Management Accounting Research*, 18(4), 453–475. <https://doi.org/10.1016/j.mar.2006.12.001>
- Molander, P., Nilsson, J.-E., & Schick, A. (2002). *Does anyone govern? The relationship between the government offices and the agencies in Sweden*. Centre for Business and Policy Research.
- Pierre, J. (2004). Central agencies in Sweden. A report from Utopia. In C. Pollitt, & C. Talbot (Eds.), *Unbundled government. A critical analysis of the global trend to agencies, quangos and contractualisation* (pp. 203–214). Routledge.
- Pollock, N., & D’Adderio, L. (2012). Give me a two-by-two matrix and I will create the market: Rankings, graphic visualisations and sociomateriality. *Accounting, Organizations and Society*, 37(8), 565–586. <https://doi.org/10.1016/j.aos.2012.06.004>
- Pollock, N., D’Adderio, L., Williams, R., & Leforestier, L. (2018). Conforming or transforming? How organizations respond to multiple rankings. *Accounting, Organizations and Society*, 64, 55–68. <https://doi.org/10.1016/j.aos.2017.11.003>
- Porter, T. S. (1995). *Trust in numbers. The pursuit of objectivity in science and public life*. Princeton University Press.
- Power, M. (2015). How accounting begins: Object formation and the accretion of infrastructure. *Accounting, Organizations and Society*, 47, 43–55. <https://doi.org/10.1016/j.aos.2015.10.005>
- Power, M. (2019). Infrastructures of traceability. *Research in the Sociology of Organizations*, 62, 115–130. <https://doi.org/10.1108/S0733-558X20190000062007>
- Power, M. (2021). Modelling the microfoundations of the audit society: Organizations and the logic of the audit trail. *Academy of Management Review*, 46(1), 6–32. <https://doi.org/10.5465/amr.2017.0212>
- Rainelli-Weiss, H., & Huault, I. (2016). Business as usual in financial markets? The creation of incommensurables as institutional maintenance work. *Organization Studies*, 37(7), 991–1015. <https://doi.org/10.1177/0170840615626463>

- Reilley, J., & Scheytt, T. (2019). A calculative infrastructure in the making: The emergence of a multi-layered complex for governing healthcare. *Research in the Sociology of Organizations*, 62, 43–68. <https://doi.org/10.1108/S0733-558X20190000062004>
- Rindova, V. P., Martins, L. L., Srinivas, S. B., & Chandler, D. (2018). The good, the bad, and the ugly of organizational rankings: A multidisciplinary review of the literature and directions for future research. *Journal of Management*, 44(6), 2175–2208. <https://doi.org/10.1177/0149206317741962>
- Ringel, L., Espeland, W., Sauder, M., & Werron, T. (2021). Worlds of rankings. *Research in the Sociology of Organizations*, 74, 1–23.
- Samioli, R. (2012). Commensuration and styles of reasoning: Venice, cost-benefit, and the defence of place. *Accounting, Organizations and Society*, 37(4), 382–402. <https://doi.org/10.1016/j.aos.2012.04.001>
- Sauder, M., Chun, H., & Espeland, W. (2021). The garbage can model and organizational metrics. *Research in the Sociology of Organizations*, 74, 175–197. <https://doi.org/10.1108/S0733-558X20210000074033>
- Sauder, M., & Espeland, W. N. (2006). Strength in numbers? The advantages of multiple rankings. *Indiana Law Journal*, 81(1), 205–217.
- Sauder, M., & Espeland, W. N. (2009). The discipline of rankings: Tight coupling and organizational change. *American Sociological Review*, 74(1), 63–82. <https://doi.org/10.1177/000312240907400104>
- Scott, S., & Orlikowski, W. (2012). Reconfiguring relations of accountability: Materialization of social media in the travel sector. *Accounting, Organizations and Society*, 37(1), 26–40. <https://doi.org/10.1016/j.aos.2011.11.005>
- Slager, R., & Gond, J.-P. (2022). The politics of reactivity: Ambivalence in corporate responses to corporate social responsibility ratings. *Organization Studies*, 43(1), 59–80. <https://doi.org/10.1177/0170840620964980>
- Stark, D. (2020). The performance complex. In D. Stark (Ed.), *The performance complex: Competitions and valuations in social life* (pp. 1–30). Oxford University Press.
- Sundström, G. (2003). *Stat på villovägar. Resultatstyrningens framväxt i historiskt-institutionellt perspektiv (The state at a loss. The development of management by objectives from a historical-institutional perspective)*. Stockholm University.
- Svårdsten, F. (2015). In the absence of detailed steering: A governmental attempt to address the issues of recentralization and detailed performance control. *Scandinavian Journal of Public Administration*, 19(1), 109–127.
- Svårdsten, F. (2019). The ‘front stage’ of substance auditing: A study of how substance auditing is presented in performance audit reports. *Financial Accountability & Management*, 35(2), 199–211. <https://doi.org/10.1111/faam.12190>
- Swedish Government. (2007). *Att styra staten – Regeringens styrning av sin förvaltning*. (Governing the state – the Government’s governance of agencies) (SOU 2007:75).
- Swedish National Financial Management Authority. (2009). *Handledning resultatredovisning*. (Guidelines for performance reporting).
- Swedish National Financial Management Authority. (2012). *Prestationer, volymer och kostnader. Utvärdering avreglerna för resultatstyrningen*. (Outputs, volumes and costs. Evaluation of the rules for managing for results). (ESV 2012:27).
- Swedish Parliament. (2009). *Offentlig förvaltning för demokrati, delaktighet och tillväxt (Public administration for democracy, participation and growth)*. (Prop. 2009/10:175).
- Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research. *Sociological Theory*, 30(1), 167–186. <https://doi.org/10.1177/0735275112457914>
- Unerman, J., Bebbington, J., & O’Dwyer, B. (2018). Corporate reporting and accounting for externalities. *Accounting and Business Research*, 48(5), 497–522. <https://doi.org/10.1080/00014788.2018.1470155>