Public Management Review

ISSN: 1471-9037 (Print) 1471-9045 (Online) Journal homepage: https://www.tandfonline.com/loi/rpxm20

Constitutive Effects of Performance Indicators: Getting beyond unintended consequences

Peter Dahler-Larsen

To cite this article: Peter Dahler-Larsen (2014) Constitutive Effects of Performance Indicators: Getting beyond unintended consequences, Public Management Review, 16:7, 969-986, DOI: 10.1080/14719037.2013.770058

To link to this article: https://doi.org/10.1080/14719037.2013.770058

Published online: 20 Apr 2013.

Submit your article to this journal

Article views: 2991

View Crossmark data

Citing articles: 67 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rpxm20

Public Management Review, 2014
Vol. 16, No. 7, 969–986, http://dx.doi.org/10.1080/14719037.2013.770058

Abstract

The idea that performance indicators in public management have unintended consequences is almost as old as performance measurement itself. But, is 'unintended consequences' an appropriate and insightful idea? The very term rests on an identification of intentions and assumptions about validity that are demonstrably problematic. Based on a distinction between trivial and advanced measure fixation, an argument is made for constitutive effects that are based on less problematic assumptions. Through this conceptual move, the political dimension of performance indicators is appreciated. The conceptual dimensions of constitutive effects are carved out, empirical illustrations of their applicability are offered and implications discussed.

Key words

Performance indicators, unintended effects, constitutive effects, performance paradox

CONSTITUTIVE EFFECTS OF PERFORMANCE INDICATORS

Getting beyond unintended consequences

Peter Dahler-Larsen

Peter Dahler-Larsen
Department of Political Science
University of Copenhagen
Denmark
E-mail: pdl@ifs.ku.dk

Routledge
Taylor & Francis Group

© 2013 Taylor & Francis

INTRODUCTION

Performance management and indicators are not just a fashion, but also a mega-trend in public management (Røvik, 1998). As old as this mega-trend is a literature that looks at the downsides of performance indicators. Measure fixation may lead to gaming defined as a 'deliberate subversion' to 'hit the target and miss the point' (Bevan and Hood, 2006). Closely related phenomena are tunnel vision, cream skimming (cherry picking), effort substitution, suboptimization, myopia, misrepresentation, misinterpretation and ossification (Smith, 1995).

More than 50 years ago, Ridgway (1956) pointed to *dysfunctional* consequences of performance measurement, and the same term is still in use (Feller, 2002; Propper and Wilson, 2002: 265). Bouckaert and Balk (1991) describe *diseases* in indicator systems. Others suggest that there are *perverse effects* (de Bruijn, 2001; van Thiel and Leeuw, 2002; Munro, 2004). But, do derogatory terms enhance analytical understanding? Calling some effects, 'perversions' or 'diseases', for example, is based on tacit assumptions of a healthy or normal form of performance measurement that are neither carefully explained nor justified. The same is true for 'misguided', 'mistaken', 'malevolent' (Feller, 2002: 449), 'misinterpreted' behaviour, where performance indicators are 'misused' (Perrin, 1998). Since one person's use can be another person's misuse (Shulha and Cousins, 1997), the prefix 'mis-' is not extraordinarily helpful.

While 'true believers' see indicators as inherently positive, and 'active doubters' hold negative ideological views, the middle ground holds the pragmatic view that under some conditions, indicators have unintended effects, in spite of good intentions (Norman, 2002). The term unintended consequences of indicators is so frequently found in the literature (McBriarty, 1988; Smith, 1995; Doran, 2001; Vulliamy and Webb, 2001; van Thiel and Leeuw, 2002; Robinson, 2003; Courtney *et al.,* 2004; Talbot, 2005; Weingart, 2005; Werner and Asch, 2005; Brennan, 2006; Wilson *et al.,* 2006; Espeland and Sauder, 2007; Hood, 2007; Osterloh, 2010; Osterloh and Frey, 2010) that it deserves special attention.

Although sympathetic and insightful, this position suffers from a theoretical and practical dependency on a valid identification of intentions behind the indicators. The purpose of this article is to demonstrate this deficit and to offer an alternative, which I call 'constitutive effects'. The exploration of this term fits into a recent research agenda that concerns the social, practical and political ramifications of inevitable measurement of something which remains contested and difficult to measure (Noordegraaf, 2008). This research agenda helps us understand how models of New Public Management 'have formative effects on society and are thereby constitutive' (Kaboolian, 1998: 191). This article is in line with Kaboolian's view, but focuses specifically on the constitutive role of indicators.

First, mechanisms that lead to 'unintended consequences' are unpacked and the underlying assumptions are discussed. Second, a conceptual alternative – constitutive effects – will be explained theoretically and described in some detail. Third, empirical

examples are provided that illustrate the usefulness of this alternative conceptualization. On this basis, mechanisms that produce constitutive effects under various circumstances are discussed, as are implications.

My argument is not that unintended consequences of indicators do not exist. Nor do I argue that they really are intended, after all. My argument is that *intentions* as such may not always be the best standard against which to assess the consequences of performance indicators. If we dare think beyond concepts with prefixes such as 'mis-' or 'un-', we focus not on what indicators fail to do, but what they actually accomplish, and a fresh perspective on their social and political ramifications is opened. The conceptual alternative advocated here is *constitutive effects*. By 'effect' I refer to something similar to a rhetorical effect, a meaning being constructed or a practice being established. I do not claim to demonstrate causal effects.

MECHANISMS LEADING TO UNINTENDED CONSEQUENCES: TRIVIAL MEASURE FIXATION

Two alternative views of the effects of indicators are illustrated in Figures 1 and 2 For analytical purposes, their assumptions will be dissected step by step. Figure 1 represents the conventional view or 'trivial measure fixation'.

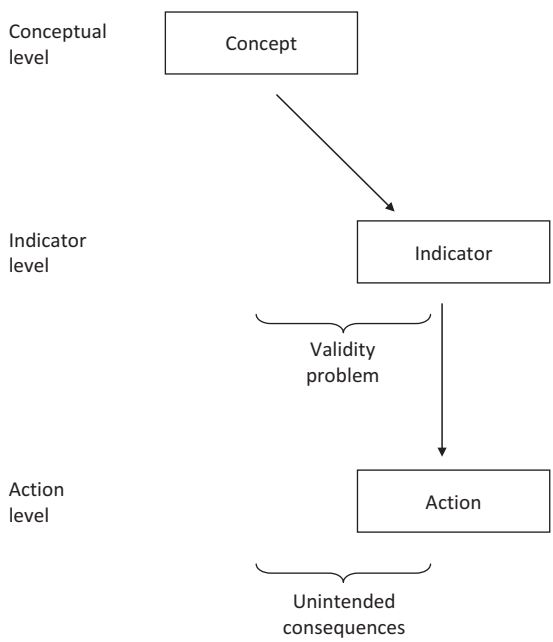

Figure 1: Trivial measure fixation

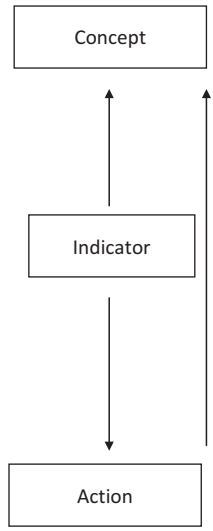

Figure 2: Advanced measure fixation

A key source behind unintended consequences is the difficulty with finding valid indicators in the public sector (van Thiel and Leeuw, 2002). Goals are multiple and sometimes vague (Baier *et al.,* 1986; Vedung, 1989; Fountain, 2001), services complex, dynamic or person oriented (Abma and Noordegraaf, 2003).

In Figure 1, the indicator is an imperfect measurement of the concept that is intended to measure. A validity problem is indicated. Nevertheless, the indicator itself guides practical action. When measure fixation (Smith, 1995) follows indicators with low validity, the conditions are present for a performance paradox (van Thiel and Leeuw, 2002), i.e. a situation where measurement of quality is poorly related to quality itself.

An example is an indicator of waiting time that counts only time spent in the waiting room and ignores all other waiting time. An unintended effect is that staff moves the patients to a chair in the hall in order to improve waiting time statistics. Perhaps staff is under many forms of pressure. But according to the view represented in Figure 1, the inappropriate indicator is responsible for unintended consequences of these pressures. In trivial measure fixation, the root of unintended consequences lies in measurement problems (van Thiel and Leeuw, 2002: 273). I call this phenomenon 'trivial measure fixation' because advocates of this view assume that it is evident and trivial for anyone that actual behaviour is a deviation from what was intended.

As illustrated in Figure 1, the two are of equal size and correspond to each other. The assumption about validity and the assumption about unintended effects are closely linked. Next, in what sense can effect be said to be 'unintended'?

PROBLEMATIC ASSUMPTIONS ABOUT UNINTENDED CONSEQUENCES

The notion of unintended consequences assumes a set of intentions against which actual consequences can be compared. These assumptions are often problematic on the following points:

1. *Identification of intentions*. If the intentions of too many stakeholders are included, one stakeholder's intended effects might correspond to the unintended effects of another. It would be unrealistic to assume that all stakeholders have similar intentions. A more narrow set of stakeholders must be assumed. More often than not in performance measurement, a principal-agent model is assumed according to which there is a lack of congruence between the goals of the agent and the goals of the principal (Smith, 1995: 283). The challenge in principal-agent theory is how to design measurement and incentive schemes that motivate the agent to perform consistent with the assumed goals of the principal. In some situations, agents are accused of 'gaming' when they literally do everything they can to make sure that what is being measured is literally being done. In that case, the researcher's tacit identification with the principal may impede impartial and in-depth analysis. In practical life, the responsibility for an indicator system may be shared between politicians, public managers and consultants, in a way that renders the notion of one principal behind the system an abstraction.
2. *Coherence of intentions*. While a clearly rational model, such as the principal-agent model, assumes that there is a coherent vision behind a performance indicator system, this idea may be unrealistic in political contexts, where there is perhaps no one agreed primary objective (Pollitt *et al.*, 2007). More than one official function, such as accountability, learning, enlightenment and information of the public (Chelimsky, 2006), may be official purposes of an indicator system. Only a murky coalition with incoherent goals can deliver the necessary majority behind some political decisions (Baier *et al.*, 1986). While gaming (a typical unintended consequence of indicators) is defined as 'hitting the target and missing the point', it is assumed that 'the point' behind political interventions is clear.
3. *Character of intentions*. A realistic perspective on political decision-making in the public sector allows for hidden agendas and concealed interests. If such motives are included, the concept of intentions is broadened, and it becomes more difficult to sustain the view that a given set of consequences are unintended. The burden of proof in an empirical demonstration of intentions becomes more difficult to carry.
4. *Timing of intentions*. Even if we accept the idea that there is a set of original intentions behind a given indicator system, what if the indicator system evolves and changes over time (Pollitt, 1989)? Could we imagine that the architect of

the system – or other stakeholders for that matter – would intentionally seek new opportunities along the way through interaction, learning processes and improvisation?

5. *Focus of analysis*. Some measurement systems are never fully implemented. Sometimes, more than one indicator system is implemented, and a given indicator may play different roles in different indicator systems. When faced not with an integrated indicator system, but a configuration of overlapping, inconsistent or half-baked indicators from different periods, how do we map *the* underlying intentions?
6. *Interest in unintended side-effects*. Especially if unintended effects are directly counter-intentional, those responsible for indicator systems would be eager to get up early in the morning to hear the latest news about the unintended effects of their inventions, if the literature on unintended consequences were correct. Architects of indicator systems are more concerned with installing new systems than with learning about the functions of existing ones (Power, 1997). Perhaps those who install performance measurement systems do not, in fact, have the intentions they are assumed to have.
7. *Empirical demonstration of intentions*. It is methodologically difficult to identify intentions. They may be unclear, ambiguous or changing over time (March and Olsen, 1976). They may not be known to those who hold them or may be reconstructed after the fact to make sense out of actions already carried out (Weick, 1979). If we lift narrow but unrealistic assumptions about intentions, we are confronted with a huge task of mapping and documenting the 'original' intentions empirically and separating them from those that do not count.

More often than not, it is just *assumed* that there was a point in time where an architect behind a given stable and bounded indicator system had a consistent and coherent vision that defined one agreed primary objective for the indicator system against which all of its consequences can be judged. Too often, it is imagined that such intention is clear, identifiable, reasonable and shared by all (Pollitt *et al.*, 2007) and, as a corollary, that the concept in which the indicator supposedly measures is clear. Without this simplistic assumption, however, the term 'unintended consequences' loses its analytical sharpness.

If assumptions about intentions are so problematic, perhaps an alternative thinking is recommendable. To build a bridge towards such thinking, we must attend to a form of measure fixation we can call advanced, not trivial.

HOW INDICATORS CAN SHAPE REALITY: ADVANCED MEASURE FIXATION

In advanced measure fixation, the indicator stands in a constitutive relation to the reality it seeks to describe. The underlying mechanism is illustrated in Figure 2. Here the indicator

is not merely a more or less valid representation of an existing concept. Instead, the indicator helps define the concept it claims to measure (an operational definition similar to how intelligence tests help define intelligence). In Figure 2, the direction of the arrow between the concept and the indicator is reversed. Again, actions may follow from the indicator even in the absence of any specific authentic concept against which the subsequent actions can be assessed. Next, the subsequent actions may also help define that construct. When people react to an indicator, they may find out how to make sense of the construct that the indicator claims to measure (Power, 1997: 51).

An example: A league table is published where schools are ranked according to the average grades given to students. Some researchers argue that this criterion is an invalid indicator of school quality since it is not corrected for the socio-economic background of the students (which is known to be a strong statistical predictor of school outcomes). Researchers argue that the league table 'cannot be used' and that it will lead to increased social segregation among schools.

For the sake of the argument, let me read the league table and decide that it can, in fact very well, be used to select a school for my children. I think that good raw average grades make up a good indicator of school quality, because of the likelihood that my children will get good grades, too, regardless of whether the good grades are caused by the socio-economic background of the other parents or by the contribution ('added value') of the school itself. I define school quality by means of the indicator, even if others may think that the indicator has low validity (thus the arrow from the indicator to the concept in Figure 2).

The league table represents and enhances a particular view of school quality. Through the actions of parents like me, and through the actions of principals and teachers who seek to enhance their ranking on the league table and attract children with parents like me, the promoted view of school quality is further confirmed (the long vertical arrow in Figure 2).

To declare all these actions, 'unintended' school quality is defined through the actions and interactions that follow the publication of the indicators. The measure fixation here is advanced rather than trivial because it is emergent and not just a deviation from a well-defined, evident and already agreed-upon concept of school quality. If we bracket our normative reservation that there *ought* to be validity, how could we conceptually understand the effects described in Figure 2?

HOW CONSTITUTIVE EFFECTS ARE PRODUCED

In Figure 2, the indicator helps define a particular version of concept such as school quality which is otherwise likely to remain messy and contested (Gallie, 1955). I shall call these effects constitutive effects. 'A constitutive rule says that something will count as something in a particular context' (Swidler, 2001: 89).

Indicators are a form of 'assisted sense-making' (Mark *et al.*, 2000) that offer interpretive keys which draw attention, define discourse and orient actions in certain directions. They incorporate a particular definition and representation of the problem they are meant to deal with. They are policy instruments which have their own structuring force upon the problems they are meant to handle and upon the people who are enrolled in doing so (Lascoumes and Le Gales, 2007). Effects of indicators are truly political as they define categories that are collectively significant in a society. Three specific mechanisms may contribute to this process.

First, indicators are produced in organizational procedures. Standard operating procedures make it possible for organizations to handle ambiguity. Through procedures for selection, quantification, retention, operations on data and forms of presentation within certain time frames, indicators produce a more or less orderly view of an otherwise disorderly reality.

Second, indicators provide a language. Along with quantification comes taxonomy that fundamentally determines what is to be counted as what. 'Activities', 'inputs', 'processes', 'outputs' and 'outcomes' are carved out of a larger, complex and dynamic reality. The 'thingness' inherent in the objects measured is often not given, but contingent upon a number of operations in place to secure standardization and objective, trustworthy measurement (Porter, 1994). Sometimes, a concept is replaced by a formula (Porter, 1994; Desrosieres, 1998: 398). In doing so, phenomena which are otherwise polyvalent and rich in meaning, are subdued to measurement along with one or a few dimensions (Porter, 1994: 396; Munro, 2004). There is a performative and constitutive aspect in how things are named as they are measured.

Third, institutional 'lock-in' takes place when incentives are connected to indicators and when those who do not attend to their score on indicators are sanctioned or ridiculed (Osterloh and Frey, 2010). If they do not score well, the energy and time they have spent appear to be wasted, as the result of their work is made invisible. The more a particular measurement formula is objectified, the less visible and less 'real' become the phenomena that are not captured by the indicator (Espeland and Sauder, 2007: 18). After some time, self-fulfilling prophecies emerge, for example, if ranking is based on reputation and reputation on ranking.

The actual effect of these mechanisms depends in practice, of course, on contextual factors and the reactions of stakeholders. Rather than advocating specific predictions, the mechanisms will now be illuminated through empirical examples. In order to qualify the concept of constitutive effects of indicators, it is necessary to add empirical flesh and blood to the three mechanisms presented and to carve out domains in which constitutive effects may occur.

DOMAINS OF CONSTITUTIVE EFFECTS: EMPIRICAL ILLUSTRATIONS

In this section, the following five domains of constitutive effects will be explored:

- Indicators define interpretive frames and world views
- Indicators define content
- Indicators define time frames
- Indicators define social relations and identities
- Indicators change their meaning as a result of their use as indicators

Empirical illustrations of these conceptual domains will now be drawn from two studies of indicator systems in Denmark. The first is a comprehensive 10-year-old testing system for adult immigrants having attended programmes in Danish as a second language. Passed tests add up to an indicator that determines the financing of language schools. The second is a new bibliometrical indicator system which helps redistribute financial resources among research institutions according to the number of publications with different points defined by an authoritative ministerial list of publishers and journals. The first case study was covered through interviews with teachers, testers and a school principal (Dahler-Larsen, 2012b), while the second hinged on participant observation, studies of official documents, including a ministerial website and an official evaluation, as well as the writing of a debated scholarly article (Dahler-Larsen, 2012a). The qualitative data from the two case studies were then sorted into the five domains described earlier (following section) and related to the three mechanisms (subsequent section).

The methodological approach is conceptually exploratory rather than hypothesis testing and comparative. The purpose is to check whether the five domains described earlier are good heuristic hooks upon which to hang empirical detail.

Indicators define interpretive frames and world views

The indicator is embedded in a larger world view, a network of interpretations, in which its definition of activity makes sense. For example, an indicator which emphasizes consumer satisfaction is an effective way to import market principles into the public sector, despite problematic effects on civic virtues (Fountain, 2001: 71). The opposite is true for indicators which focus on productivity measures without looking at consumers.

In the Danish for immigrants example, the sheer number of students passing tests constitutes an indicator that exclusively regulates income for schools. It is not emphasized how, say, teachers can build relations with students that help students manage the

many daily-life problems that immigrants encounter. Before the test system occurred, these elements of the teaching practice did play a role, according to teachers. The test system portrays teaching of a language as a goal-directed and instrumental activity. As the socio-demographic composition of immigrants change as a result of official Danish immigration policies (towards more immigrants from Western countries with better education and fewer without alphabetization), schools find that money which rewards that activity flows more generously. The world view suggested by the indicator is thus socio-economically and ethnically consequential.

In the bibliometry example, scientific activity is defined exclusively in terms of publications. By focusing on research 'output', a view is enhanced where scientific activity is similar to industrial production. Scientific inquiry, reading, collecting data, serving as an editor or reviewer, giving advice or engaging in debate are not counted.

Indicators define content

Through indicators, practices which are otherwise more intuitive and implicit are exposed to explicit criteria of success (Munro, 2004: 1082). Indicators define what is central in work. They help determine what actors should strive to accomplish in a given activity. In the Danish for immigrants example, teachers said that tests only focused on selected aspects of the use of Danish language. While some test items were school oriented and not of much practical use, leading to a 'parrot language', important learning outcomes such as pronunciation, vocabulary and background knowledge of Danish culture and society were not tested, said teachers. They acknowledged the link between test results and money streams as a fact. As a consequence, 'tests guide more than 50% of our teaching', says one teacher. There is sometimes more focus on test format than content. At the same time, some teachers say that tests are 'not completely crazy' and tests help them structure their curriculum.

In the bibliometry example, indicators that count research publications must be based on a definition. In social science, books on policy issues may contribute to a public debate and be used in teaching. So are they not research publications? Some observers predict that researchers will slice their work up in its smallest publishable parts and focus on research questions that are not risky. The number of points given to Danish or English publications has also become politically controversial. The issue is not only whether Danish taxpayers deserve Danish publications, but also how language, quality and content are related.

Indicators define time frames

Some academics fear that the increased focus on counting publications on an ongoing basis leads to increased pressure to publish quickly. Time is not innocent, however. It is

feared that time pressure inhibits deep, innovative and risky research projects. On the other hand, advocates of the bibliometrical system argue that researchers, who used to keep lengthy publications in their drawers for too long, will actually benefit from a bit of time pressure.

In the Danish for immigrants example, timing is institutionally structured, because particular tests must be passed in order to achieve a residence permit or, ultimately, citizenship. Teachers decide – in cooperation with each student – when the time is right for a particular test. Although no fixed scheme is imposed, formal and informal information from the school administration secures that teachers are attentive to time. While several critical views against time pressure were voiced among teachers, before the test system, some students could 'study forever', according to one teacher. But today there simply 'is no time' for those aspects of teaching which are not tested. The indicator system helps install time frames that teachers experience as institutionalized and therefore factual reality.

Indicators define social relations and identities

When the test system was introduced, there were heated debates among teachers of Danish. After some time, however, the debate died out, as everybody recognized that tests had come to stay. Furthermore, in times of economic crisis for schools, those who were very critical were reportedly more likely to leave than others. Teachers of Danish have developed a sense of professionalism over the years. To them, professionalism includes the skills of teaching, a not-too-personal relation to students and their many problems and an ability to handle the pressures of the test system as a part of daily life. 'If you are good, you can make it work', says one.

At the same time, the test system paves the way for a new vocabulary describing students. A conventional distinction separates weak from strong students. Weak students are those who lack literacy, who come from cultures distant from the Danish one, who have little schooling or whose memory and learning abilities have been negatively affected by torture or trauma. The test system helps promote a slightly different distinction, i.e. the distinction between fast and slow students. Fast are those who quickly pass tests. Slow are those who need more work before they can do so. Slow does not necessarily imply attention to all the complicated characteristics of the weak. Thus, the vocabulary of the indicator system supports a more one-dimensional view of the students.

One of the most immediate consequences of the new bibliometrical indicator is the heated debates it creates among academic staff. Two of the most important distinctions that separate the different interpretations among academic colleagues are optimism versus pessimism and clarity versus ambiguity. While optimists view life as easier with objective measurement, pessimists feel the new indicator adds to existing pressures on academic staff. Those with clarity guiding their interpretations believe the new indicator

is consistent with academic policies and managerial initiatives in general, but those with an ambiguous cultural view find inconsistencies and half-baked guidelines everywhere. For example, indicator points are not everywhere officially counted on the individual level, at least for the time being (although computer programs supporting the bibliometry make individual productivity figures easily available). Points equivalent to a doctoral thesis can be earned with many, easy, low-ranking publications. One research grant can override all financial effects of the publications of more than a life-time. With academic recognition or administrative responsibilities, it is possible to liberate oneself from the publication pressure.

The divisions among academics do not go away easily. Groups seek to convince others that their interpretation is correct. Especially the clear pessimists are likely to find life miserable at the university. Whether they leave the university in the long run, or whether they gradually become ambiguous pessimists, not to mention optimists, remains to be seen. The bibliometrical indicator helps raise the tensions between groups with these different interpretations. People know who they are through the figures that describe them. Statistics help define populations and collective identities (Porter, 1994).

Indicators change their meaning as a result of their use as indicators

Official indicators are themselves social constructions reflecting the variety of practices and meanings accorded to them by different groups of participants in different social contexts (Vulliamy and Webb, 2001: 361). The integrity of an indicator may come under pressure when the indicator is used in connection with legal, financial or managerial sanctions (Campbell, 1987: 306; Shadish *et al.*, 1991; Vulliamy and Webb, 2001). In both the cases, a mindset of productivity and effectiveness is superimposed upon phenomena (tests and publications) which were otherwise under the control of professional groups.

In the Danish education case, great care is taken to protect the integrity of tests. It is forbidden to take test material out of the test room, since tests are used repeatedly over time. Two testers in a particular school take pride in assuring that tests are consistent. Teachers, not testers, inform students about the results, which relieves testers of the psychological pressures of meeting with desperate, failing students. Institutional mechanisms help protect the trustworthiness of the testing, at least in terms of procedures, if not content.

In the bibliometrical case, publications are counted and approved by librarians. Apparently, the counting procedure is objective and reliable. A committee evaluates journals based on criteria such as the scope of the journal and, importantly, its procedures for peer review. Only peer-reviewed journals count. In this respect, although bibliometry may be believed to be more quantifiable, easy to understand and objective than peer review (Osterloh and Frey, 2010), the bibliometrical indicator is in fact based on the credibility of an underlying peer review. As a consequence, the

meaning of peer review changes. To motivate contributors, editors of journals and anthologies sometimes promise that a publication is peer reviewed (meaning that publications will count) *and* that the contribution will be published. While peer review has a long tradition in academia as a much-celebrated mechanism for both formative and summative evaluation, the very decision to publish the contribution is now potentially taken out of the equation, at least for editors who are in need of contributions. The meaning and function of peer review change if the decision to publish has already been made. In that way, the bibliometrical system has a profound impact of the meaning not of the indicator itself, but on another quality assurance system with which it interacts.

MECHANISMS LEADING TO CONSTITUTIVE EFFECTS

Our first mechanism, organizational procedure structuring an otherwise messy reality, is in place in both cases. A bibliometrical measurement system as well a language testing system is made of organizational procedures. Much effort is made in both cases to make the indicator system strong, trustworthy, and reliable so that its results will be accepted. The acceptability of the bibliometrical system, however, was less clear than that of the language testing system. The former was newer and had less time to achieve success in reducing ambiguity.

The second mechanism, indicators shaping language, is well illustrated in both the cases. Indicators help define a landscape of meanings, definitions, incentives and regulations in which professionals find themselves. In academia, a more results-oriented vocabulary is promoted, as publication points are 'earned'. The same is true in the case of teaching of Danish for immigrants, where 'results' have strong socio-economic overtones because some immigrants are 'fast' while others are 'slow'. Teachers and schools are motivated to look at 'fast' (Western and well-educated) students more favourably. Hereby, recent Danish immigration policies unfold. The terminology transported by indicators invites teachers and schools to be strategic allies in the implementation of these policies.

These effects are enforced through our third mechanism, institutional lock-in. In both the cases, national legislation helps make the indicator system institutionally solid. Incentives and sanctions work strongly in the Danish education case and moderately in the bibliometry case. In the former case, incentives are almost exclusively linked to the testing system, while in the latter case, bibliometrical remuneration is just one among several sources of incomes for universities.

Incentives work at several levels of analysis. In none of the earlier cases are indicator scores broken down to individuals, but everybody knows that data are available to do so. In both the cases, discourse, rhetoric and managerial initiatives seem strong enough to make individuals realize the force of the indicator even if the individual scores are not made available. The acceptability of the indicator system among employees varies. In the Danish education case, although some teachers held very critical views, the test-

based indicator system was generally accepted as 'not completely crazy' and under all circumstances a part of existing reality.

Teachers of Danish integrate a day-to-day management of the testing system into their sense of professionalism. One of the reasons why the teachers accepted the testing system was the perceived lack of structuration of their methods and objectives in the past. In contradistinction, academic values and traditions offers an interpretive platform against which the new biblimetric indicator is evaluated negatively by many. This observation suggests that those who define professional identity and shape stories of the past influence the relative acceptability of indicators among professionals. It would be politically unsophisticated not to notice how both professional groups, teachers and academics respectively, must define themselves in interaction with the indicator systems. Constitutive effects are interactive phenomena.

When performance measurement of, say, research or Danish for immigrants is politically enacted, many ambiguities and priorities are transported to those who design the measurement system and seek to make it work in practice (de Bruijn, 2001), such as how tests are designed and what publications count. Decisions of political relevance are pushed down the 'implementation chain'.

In other words, although institutional lock-in may be an important mechanism, it is not complete and attempts to create lock-in may be followed by strategic moves at other levels of the implementation chain than where it was introduced. Different measurements may not only increase control and reduce risk for some, but decrease control and increase risk for others, so spiralling regulatory logics may follow, and the practical meaning of indicators may be contested (Giddens, 1994; Rothstein *et al.,* 2006). A promising line of research would thus be to combine an institutional approach to indicators with a dynamic and interactive stakeholder analysis (Elias *et al.*, 2002).

DISCUSSION AND CONCLUSION

In assessing the usefulness of the idea of constitutive effects, constructivists may be immediately comfortable with the idea, while rationalists and functionalists may still find value in the idea of unintended consequences. Thus, the concept hinges on paradigmatic foundations.

Alternatively, its usefulness may depend on situational circumstances. If there is consensus about intentions behind indicators in a given situation, it may be relevant to look after unintended consequences. If there is ambiguity or dissensus about such intentions in a complex and dynamic context, then language, discourse and social interaction may help produce *de facto* effects of indicators, in which case the idea of constitutive effects may be more relevant.

Constitutive effects of indicators are not occurring with the same intensity and frequency everywhere. Our analysis points to factors which may help explain such variation: The age and solidity of organizational procedures, the strength and

uniformity of incentives, and the contestability of language. Many other factors (such as managerial styles, professional power, working climates, substance of work, publication and media attention) may help explain how much constitutive effects unfold in particular contexts. The main contribution of this article, however, lies not in comparative factors, but in the very conceptualization of constitutive effects and in pointing out the mechanisms and the domains where they operate.

Perhaps the strongest political effect of indicators is how they constitute a reality that is put on stage so that it can be acted upon (Desrosiéres, 1998: 352), and thus how they define a strategic landscape in which practitioners must navigate. This landscape is political, as it consists of policy-related categories by means of which society seeks to manage itself and thereby represents itself and its values (Vestman and Conner, 2006; Rosanvallon, 2009). Perhaps we have here found a central ingredient in how the instruments of New Public Management are socially and politically constitutive (Kaboolian, 1998: 190).

Paradoxically, instruments to enhance visibility and transparency, such as indicators, are themselves interpretations that set their own political effects in motion. Sometimes, these effects are self-fulfilling prophecies, such as when ranking and performance reinforce each other. Indicators become the way through which the world is defined. When an indicator operates, it may also influence the operation of other forms of checking and defining quality. As illustrated in our case material, bibliometrical measures influenced the definition of peer review. The constitutive effects of such 'finding' are not just negative epiphenomena, but also contributions to an interrelated network of interpretations.

If an evaluation subsequently finds that the bibliometrical system has worked efficiently because it has helped increase the number of publications over time; then that way of understanding the impact of indicators is already based on the world view implemented by the indicator. The constitutive effect of having defined research only in terms of its outputs should be attended to rather than taken for granted. Pre- and post-tests that measure changes in a defined indicator should not be isolated from the (contested) institutionalization of world views connected to that indicator. The observation that one form of indicator influences subsequent forms of knowledge creation would suggest that political effects of indicators may be far-reaching.

Constitutive effects are a sensitizing concept which should be filled with further empirical material. The domains listed earlier (world view, content, time frame and relations) have proved useful as a conceptual scheme to follow up on these effects. If controversy and disagreement are acknowledged, there is a limit to how precisely constitutive effects can be predicted, but their interactive and contested nature can be attended to. If indicators play a role in the reconfiguration of society, consider their contestation as normal and good (Pestre, 2003: 260).

REFERENCES

Abma, T. and Noordegraaf, M. (2003) Public Mangers Amidst Ambiguity: Towards a Typology of Evaluation Practices in Public Management. *Evaluation*, 9:3 pp285–306.

Baier, V., March, J. G. and Sætren, H. (1986) Implementation and Ambiguity. *Scandinavian Journal of Management Studies*, 2 pp197–212.

Bevan, G. and Hood, C. (2006) What's Measured Is What Matters: Targets and Gaming in the English Public Health Care System. *Public Administration*, 84:3 pp517–38.

Bouckaert, G. and Balk, W. (1991) Public Productivity Measurement: Diseases and Cures. *Public Productivity & Management Review*, 15:2 pp229–35.

Brennan, R. ed. (2006) 'Perspectives on the Evolution and Future of Educational Measurement' in *Educational Measurement* 1–16. Westport, CT: Greenwood Publishing.

Campbell, D. T. (1987) Guidelines for Monitoring the Scientific Competence of Preventive Intervention Research Centers: An Exercise in the Sociology of Scientific Validity. *Knowledge: Creation, Diffusion, Utilization*, 8 pp389–430.

Chelimsky, E. (2006) 'The Purposes of Evaluation in a Democratic Society' in I. Shaw, J. Greene, and M. M. Mark (eds) *The Sage Handbook of Evaluation* 33–55. London: Sage Publications.

Courtney, M. E., Needell, B. and Wulczyn, F. (2004) Unintended Consequences of the Push for Accountability: The Case of National Child Welfare Performance Standards. *Children and Youth Services Review*, 26 pp1141–54.

Dahler-Larsen, P. (2012a) 'Betydningen Af Bibliometriske Indikatorer. De Konstitutive Virkninger Af Forskningsevaluering' in F. Jan and B. P. David (eds) *Hvad Er Videnspolitik? Institutioner, Ledelse Og Styring* 337–58. Copenhagen: Nyt fra Samfundsvidenskaberne.

Dahler-Larsen, P. (2012b) Constitutive Effects as a Social Accomplishment: A Qualitative Study of the Political in Testing. *Education Inquiry*, 3:2 pp171–86.

De Bruijn, H. (2001) *Managing Performance in the Public Sector*, New York: Routledge.

Desrosiéres, A. (1998) *The Politics of Large Numbers: A History of Statistical Reasoning*, Cambridge, MA: Harvard University Press.

Doran, H. (2001) *Dissertation: Evaluating the Consequential Aspect of Validity on the Arizona Instrument to Measure Standards*, Tucson, AZ: University of Arizona.

Elias, A. A., Cavana, R. Y. and Jackson, L. S. (2002) Stakeholder Analysis for R&D Project Management. *R&D Management*, 32:4 pp301–10.

Espeland, W. and Sauder, M. (2007) Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, 113:1 pp1–40.

Feller, I. (2002) Performance Measurement Redux. *American Journal of Evaluation*, 23:4 pp435–52.

Fountain, J. E. (2001) Paradoxes of Public Sector Customer Service. *Governance*, 14:1 pp55–73.

Gallie, W. B. (1955) Essentially Contested Concepts. *Proceedings of the Aristotelian Society*, 56 pp167–98.

Giddens, A. (1994) *Modernitetens Konsekvenser*, København: Reitzel.

Hood, C. (2007) Public Service Management by Numbers: Why Does It Vary? Where Has It Come From? What Are the Gaps and the Puzzles? *Public Money and Management*, 27:2 pp95–102.

Kaboolian, L. (1998) The New Public Management: Challenging the Boundaries of the Management vs. Administration Debate. *Public Administration Review*, 58:3 pp189–93.

Lascoumes, P. and Le Gales, P. (2007) Introduction: Understanding Public Policy Through Its Instruments – From Nature of Instruments to the Sociology of Public Policy Instrumentation. *Governance*, 20:1 pp1–21.

March, J. G. and Olsen, J. P. (1976) *Ambiguity and Choice in Organizations*, Bergen: Universitetsforlaget.

Mark, M. M., Henry, G. and Julnes, G. (2000) *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Public and Nonprofit Policies and Programs,* San Francisco, CA: Jossey-Bass.

McBriarty, M. A. (1988) Performance Appraisal: Some Unintended Consequences. *Public Personnel Management*, 17 p4.

Munro, E. (2004) The Impact of Audit on Social Work Practice. *British Journal of Social Work*, 34 pp1075–95.

Noordegraaf, M. (2008) Meanings of Measurement. *Public Management Review*, 10:2 pp221–39.

Norman, R. (2002) Managing Through Measurement or Meaning? Lessons From Experience with New Zealand's Public Sector Performance Management Systems. *International Review of Administrative Sciences*, 68:4 pp619–28.

Osterloh, M. (2010) Governance by Numbers: Does It Really Work in Research? *Zeitschrift Für Sozialtheorie*, 32 pp267–83.

Osterloh, M. and Frey, B. S. (2010) *Academic Rankings Between the 'Republic of Science' and 'New Public Management'*, Working Paper. Zurich: CREMA – Center for Research in Management, Economics and the Arts.

Perrin, B. (1998) Effective Use and Misuse of Performance Measurement. *American Journal of Evaluation*, 19 pp367–79.

Pestre, D. (2003) Regimes of Knowledge Production in Society: Towards a More Political and Social Reading. *Minerva*, 41 pp245–61.

Pollitt, C. (1989) Performance Indicators in the Longer Term. *Public Money & Management*, 9 pp51–5.

Pollitt, C., Harrison, S., Bal, R., Dowswell, G. and Jerak, S. (2007) *Conceptualising the Development of Performance Measurement Systems*. Paper presented at the European Group for Public Administration Annual Conference, Madrid, September.

Porter, T. M. (1994) Making Things Quantitative. *Science in Context*, 7:3 pp389–407.

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

Propper, C. and Wilson, D. (2002) The Use and Usefulness of Performance Measures in the Public Sector. *Oxford Review of Economic Policy*, 19 p2.

Ridgway, V. F. (1956) Dysfunctional Consequences of Performance Measurements. *Administrative Science Quarterly*, 1:2 pp240–7.

Robinson, P. (2003) Government Accountability and Performance Measurement. *Critical Perspectives on Accounting*, 14 pp171–86.

Rosanvallon, P. (2009) *Demokratin Som Problem*, Hägersten: Tankekraft Förlag.

Rothstein, H., Huber, M. and Gaskell, G. (2006) A Theory of Risk Colonization: The Spiralling Regulatory Logics of Societal and Institutional Risk. *Economy and Society*, 35:1 pp91–112.

Røvik, K. A. (1998) *Moderne Organisasjoner. Trender I Organisasjonstenkingen Ved Tusenårsskiftet*, Bergen-Sandviken: Fagbokforlaget.

Shadish, W. R., Cook, T. D. and Leviton, L. C. (1991) 'Social Program Evaluation: Its History, Tasks and Theory' in W. Shadish, T. Cook and L. Leviton (eds) *Foundations of Program Evaluations* 19–35. Newbury Park, CA: Sage.

Shulha, L. M. and Cousins, J. B. (1997) Evaluation Use: Theory, Research, and Practice Since 1986. *Evaluation Practice*, 18:3 pp195–208.

Smith, P. (1995) On the Unintended Consequences of Publishing Performance Data in the Public Sector. *International Journal of Public Administration*, 18:2&3 pp277–310.

Swidler, A. (2001) 'What Anchors Cultural Practices' in T. R. Schatzki, K. K. Cetine, and E. V. Savigny (eds) *The Practice Turn in Contemporary Theory* 74–92. London: Routledge.

Talbot, C. (2005) 'Performance Management' in F. Ewan, L. E. Lynn and C. Pollitt (eds) *The Oxford Handbook of Public Management* 491–517. Oxford: Oxford University Press.

van Thiel, S. and Leeuw, F. (2002) The Performance Paradox in the Public Sector. *Public Performance and Management Review*, 25:3 pp267–81.

Vedung, E. (1989) *Utvärdering I Politik Och Forvaltning*, Lund: Studentlitteratur.

Vestman, O. K. and Conner, R. F. (2006) 'The Relationship Between Evaluation and Politics' in Shaw, I. F., Greene, J. C. and Mark, M. M. (eds) *The Sage Handbook of Evaluation* 225–42. New York: Sage Publications.

Vulliamy, G. and Webb, R. (2001) The Social Construction of School Exclusion Rates: Implications for Evaluation Methodology. *Educational Studies*, 27:3 pp357–70.

Weick, K. E. (1979) *The Social Psychology of Organizing*, Reading, MA: Addison-Wesley.

Weingart, P. (2005) Impact of Bibliometrics Upon the Science System: Inadvertent Consequences? *Scientometrics*, 62:1 pp117–31.

Werner, R. and Asch, D. (2005) The Unintended Consequences of Publicly Reporting Quality Information. *JAMA*, 293:10 pp1239–44.

Wilson, D., Crox, B. and Atkinson, A. (2006) 'What Gets Measured Gets Done': Headtearchers Responses to the English Secondary School Performance Management System. *Policy Studies*, 27:2 pp153–71.