

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

Abstract

Given ongoing changes in the social structuration of evaluation, the circulation of evaluative information is becoming an increasingly important factor in the production of constitutive effects of evaluation. There is power in the circulation of evaluative information, and therefore also power in attempts to regulate or curb it. Based on one case exemplifying massive circulation of evaluative information (Google Scholar), and one with very limited circulation (workplace assessments), we show how infrastructure, interests and institutional rules are useful analytical entry points into studies of circulation. The paper proposes a new theoretical and empirical attention to circulation of evaluative information as a fragile and contested phenomenon.

Introduction

There was a time when the person who controlled access to a physical evaluation report also controlled access to the evaluative information found in the report. There was also a time when a frequently cited book claimed that the best way to secure utilization of evaluation was to improve the quality of the interaction between the evaluator and key primary users and their personal relations (Patton 1978).

Today, the social structuration of evaluation has changed. In many large-scale evaluation systems (from the OECD's PISA results to Tripadvisor and IMDb), it is impossible to identify a specific evaluator and a specific end-user, with the evaluative information itself usually being produced digitally and globally (Merry, Davis and Kingsbury 2015).

In the form of a number of practices including auditing, performance management, quality assurance and benchmarking, evaluation has become such a widespread social phenomenon that the term "The Evaluation Society" (Dahler-Larsen 2012) is often referred to. This is a society with flows of evaluative information (Rist and Stame 2006) which fundamentally restructure organisational relations (Power 1997), our collective understanding of global problems (Merry, Davis and Kingsbury 2015) and lead to new forms of social inequality (Mau 2019). Evaluative information is perhaps used as the *only* basis for decisions (Roberts, 2017), thus becoming a "Tyranny of Metrics" (Mueller 2018).

At the heart of this transformation of how evaluative information is embodied in social structures and relations lies its *circulation*. Today, circulation means rapid transportation of large quantities of evaluative information through time and space supported by information technology. The bold idea advocated in this article is that the underlying *circulation* itself is an important, but somewhat overlooked, phenomenon. While the literature on evaluation has expanded the notions of use over several decades (Vedung 1997; Weiss 1977; Dahler-Larsen 2014; Espeland and Sauder 2016; Esposito and Stark 2019; Kerpershoek, Groenleer and de Bruijn 2016), it has focused less on how the social transformation of the communicative structures and relations in which

Keywords:
circulation,
diffusion,
performance indicators,
workplace assessments,
Google Scholar

evaluation is embedded conditions the performativity of evaluation (Dambrin and Robson 2011).

If there is power in evaluation, there is power in circulation of evaluation information. Power is also required to stop it. The circulation of evaluation is not random. It is more or less regulated through struggles involving infrastructures, interests and rules. In unfolding this argument, the article contributes to the understanding of circulation as a specific phenomenon which contributes to power in evaluation and the power of evaluation.

The article sets out to answer three questions:

1. How have the social conditions for the production and use of evaluative information changed in a way which makes the circulation of evaluative information particularly important?
2. What roles do infrastructure, interests and rules play in circulation of evaluative information?
3. Which power effects of evaluative information are made possible by circulation?

The purpose is not to contribute to theories of power as such. However, a few clarifying remarks about the notion of power are pertinent. A classical notion equates power with resources that A possesses and that enable A to make B act in a way which B would not otherwise have done (Thomsen 2002). Modern conceptions of power view power as more discursive, more relational, more dispersed, more productive. Power is distinct from mere coercion. Power includes self-discipline. Power resides, for example, in defining a regime of truth and a strategic institutional landscape in which actors come to see themselves and their options for action.

That said, both the 'production' and the 'use' of evaluative information are already 'entangled' in power at the same time as they produce 'power effects'. Perhaps power is not just an isolated and independent 'cause' of a chain of events. Perhaps it is better understood as an ever-present aspect of how social relations are structured on a dynamic basis. Circulation of evaluative information is consequently both a result of the distribution of power in social relations, and a source of power that can be used in attempts to restructure these relations.

The article proceeds by addressing questions 1, 2 and 3 in a logical order. Space constraints allow for just a brief narrative account of two illustrative cases. They are well chosen. One case, Google Scholar, comprises an abundance of circulation, the other (workplace assessments or WPAs), very limited circulation, but it is an emerging issue in debates. The two cases are subsequently analysed using infrastructures, incentives and rules as analytical categories. A dialogue between theoretical and empirical perspectives will take place en route. The article ends with a conclusion and recommendations for future studies.

Changes in the Production of Evaluative Information That Make Circulation Particularly Important

Highlighting contemporary changes in the social forms and shape of evaluation requires an analytical yardstick. To this end, we use a classical notion of

evaluation, where evaluation is a research-based, stand-alone inquiry into a specific problem. It is requested by an authority legitimately situated in the parliamentary chain of command (Vedung 1997). Its quality is judged in terms of its application of appropriate valid and reliable methods, its clearly defined and pre-established criteria, its logical conclusions, and its intended, subsequent use in political or managerial decision-making. Using this ideal type as a contrast, we observe four changes in the social configuration of evaluation:

i) a shift in emphasis from government to governance meaning that evaluation, its criteria and its use are no longer solely defined on the basis of the parliamentary chain of control in a representative democracy. Evaluation is not only used as an instrument in the analysis of policies. It is instead an integrated part of policies (Peters and Pierre 2020), especially in hard New Public Management (NPM) regimes focusing on performance indicators and contracts with public and private providers of service. These regimes have a number of unintended side effects beyond the control of central government.

In addition, for example in softer NPM regimes, the mandate to evaluate is sometimes delegated to local networks and configurations of authorities, consulting companies, users of services, citizens, and others involved in innovation, collaboration, or co-production. While alternative forms of evaluation sometimes, but not always, subscribe to alternative definitions of democracy (such as participatory or deliberative democracy), we are now in a situation where evaluation often operates without any commitment to any notion of democracy (Dahler-Larsen 2019). Whether evaluation is recycled, incorporated in, or in dialogue with democratic structures is becoming an empirical question that cannot be taken for granted.

Many providers of evaluative information are not under parliamentary control (Brown 2017). These include credit rating agencies, consulting companies with their particular set of performance indicators, test providers, accreditation agencies, international organizations producing their own indicators corresponding to various problem-definitions (Davis et al. 2012), and private companies producing consumer-oriented websites. Thus, a shift from government to governance takes many forms and have different normative underpinnings.

ii) A shift from ad-hoc evaluation-to-evaluation systems (Rist and Stame 2006; Leeuw and Furubo 2008). When evaluation takes the form of evaluation systems, continuous flows of evaluative information are being produced. These streams of information are often in interaction with management systems (such as risk management, quality assurance, performance management, and accreditation) (Power 2016).

iii) a shift in epistemic foci. Evaluation systems are often designed to incorporate a particular world view and epistemic focus (Leeuw and Furubo 2008). In public management and governance, Peters and Pierre (2020) observe a shift from evaluation of programs and policies to audits of the administration of policies. This change embodies a change in evaluative questions from questions about policy designs and effects to questions about administrative procedures and the ability of institutions to live up to particular performance criteria. As a

consequence, there is less emphasis on whether a given policy works and more emphasis on performance criteria.

iv) Next, the status of validity and reliability have changed from being seen as logical preconditions of evaluative information to being contextual factors of variable importance. In the classical notion of evaluation, sometimes even referred to as 'evaluation research', the transparency of methods and criteria is unquestionably understood as a precondition for the legitimacy and quality of an evaluation. However, empirical studies find that the perceived methodological quality of an evaluation is one factor among many which influence the use of evaluation, but only under particular conditions. For example, if evaluation findings do not confirm what people usually believe, they subject an evaluation to a truth test which hinges on the credibility of the methods used in the evaluation (Weiss 1977). Ledermann (2012) adds that if the pressure for action and change is strong in a given situation, that may help reduce the importance of such truth tests.

With abstract values like sustainability, equality, quality (Dahler-Larsen 2019b) and transparency (Bernstein 2017), "risky" operationalizations are necessary (Power 2019). In other words, as the thematic scope of evaluation expands, classical methodological requirements are relaxed over time. Even more radically, Hesselmann and Schendzielorz (2019) argue that once certain indicators are travelling across time and place, it becomes debatable what they actually measure. Sometimes the indicator helps define the concept it claims to measure. This reversal places the original role of "validity" on its head and reinforces constitutive effects of measurement (Dahler-Larsen 2014).

In recent years, we have seen a diffusion of evaluative information where there is no or limited account of its methodological background and underpinnings. The paradoxical nature of transparency is becoming clear: Although evaluative information is usually produced in order to enhance transparency, the production of evaluative information itself is often not very transparent (Espeland and Sauder 2016: 3). Large international organizations produce evaluative information in ways which are often impossible to access or understand for ordinary citizens. The barriers include lack of visibility regarding who is actually responsible for the production of the information; distance in time and place regarding institutional channels for insights and influence; the complicated statistical and expertise-based vocabulary in which the production of evaluative information is cast; the different rhythms in time between fast distribution of information and slow democratic processes (Rosa 2014); the lack of willingness to share the exact conditions under which evaluative information is produced because diffusion of this knowledge would make it easy to influence some scores in a manipulative way.

Admittedly, not all evaluation is subject to all these shifts to the same degree; and not all of these shifts lead to the same results. However, together they lend support to a relevant theoretical proposition: If evaluation becomes more systematized and institutionalized, but at the same time more or less liberated from direct democratic control, and less dependent on the kind of legitimacy that could be gained from methodological and democratic transparency, then circulation of evaluative information becomes even more

critical as a source of power and also a matter of potential controversy. The present circulation of data is simply taking place so fast that it runs ahead of the methodological and democratic concerns that were conventionally put in place to secure the credibility and social acceptability of particular forms of data collection. Democratic and deliberate processes are simply too slow to catch up (Rosa 2014), but that does not inhibit the present form of circulated evaluative information from functioning in practice for some users (Esposito and Stark 2019).

What Roles do Infrastructure, Interests and Rules Play in Circulation of Evaluative Information?

On the road towards a more specific analysis of circulation of evaluative information, this section provides an argument for a focus on three factors: The infrastructure which makes circulation possible; the interests and incentives connected to various roles such as ‘object’, ‘provider’, and ‘user’ of evaluative information; and the institutional rules which regulate circulation. These factors are interrelated, and part of a large “hinterland” (Law 2004) which makes circulation of a particular kind of knowledge possible. Nevertheless, as the subsequent case stories will show, these three factors can serve as useful analytical entry points.

Infrastructure

By infrastructure we mean not only the technical means of transportation of data (such as computers, networks, websites), but also the expertise, the organizational capacity to establish what Latour (1987) calls “centres of calculation”, and the construction of networks, alliances and institutional arrangements necessary for large-scale production and diffusion of data. Digital information, the internet, and its users, provide an extremely powerful alliance, but it is sometimes forgotten how the operation of that alliance presupposes that all information be suppressed into particular digital forms.

A conspicuous example of the role of infrastructure is international PISA results, which require an international organization with access to schools in many countries. Consent from hundreds of gatekeepers is needed. Tests must be developed which are translated and calibrated. Expertise must be mobilized and brought in to develop and defend the methodology and the results. Without computers and commensuration of standardized testing results across languages and cultures, it would all be impossible. However, once the comprehensive configuration of infrastructural elements needed to operate this machinery is established, it is virtually impossible to promote a competitive alternative capable of delivering a similar amount of internationally comparative data. The epistemic regime established through PISA can become dominant only because the infrastructural configuration necessary to support it is in place.

Interests

Yet, in practical life, circulation of such data does not happen by itself. It would be costly for any centre of calculation if it were to force human subjects to provide data and force users to use the data. It is more advantageous to create

associations, networks, connections and alliances (Latour 2005) where people are interested in transporting and using data. For example, journalists have been very interested in reporting about PISA results.

Sometimes an overlap of roles is created so that users are also providers. An example is found in consumer reviews, where people can find the perspectives of “people like me” (Esposito and Stark 2019). Another example is in performance data where people who have relatively good scores are interesting in enhancing the use of the system which provides these data (Fochler and de Rijcke 2017).

In these situations, an alignment of interests further enhances the circulation of a given performance information system.

However, circulation of evaluative information is not always in everybody’s interest. Espeland and Sauder (2016) describe how prospective law students in the US use rankings of law schools. Since students actually use these rankings as a quick way to tame the seemingly limitless information that is available about these schools, then admission officers, administrators, alumni, and managers have to pay attention to these data, too (Espeland and Sauder 2016: 59). Media and employers also put emphasis on the status of the school which is bestowed by rankings in no small measure (Espeland and Sauder 2016: 107). Although some deans and some faculty see this effect as highly problematic, the coalition of interests among students, managers, media, and the employers is so strong that the rest must follow. A key part of the explanation is probably that the quality and quantity of the applicants to the school – in an increasingly competitive world – is a strong determinant of the quality of the graduates and the perceived status of the school both directly and indirectly. Therefore, when prospective students use rankings as a basis for their choice of college, a self-fulfilling prophecy is set in motion. Circulation of rankings through media make these alliances possible in the first place, but alliances also enforce the power of the rankings and creates a demand for further circulation, even if these alliances are not formed without friction.

Neither the circulation nor the use of rankings hinges solely on whether the rankings themselves are valid and reliable indicators of the quality of the law schools (however that might be defined). But over time, an increasing number of people take the rankings as a functional substitute for a measure of quality. In this social construction of the quality of law schools, circulation of the rankings plays an absolutely central role.

Rules

Public authorities can regulate the diffusion of performance produced internally in public organizations in both positive and negative ways. For example, under NPM regimes, performance data on schools, higher education institutions and hospitals are made publicly available to allow users of these services to make informed choices. There are also examples of restrictions. For example, the Danish legislation on national tests in schools require test results to be known at different levels of the parliamentary chain of control, but it is not permitted to publish test data where schools are compared. The exact design of such rules is politically decided and can be politically contested.

For example, when it was proposed in Denmark to publish average final exam results controlled for socio-economic background variables such as the proportion of single parents in the school district, some citizens found these data socially degrading and therefore resisted their publication.

There is an increasing amount of performance data about public organizations which are not under direct control by public legislation except in a very broad sense (such as protecting of anonymity, GDPR regulations, etc.). As long as a provider of data respects these very broad rules, data such as consumer reviews can be published based on very few respondents, methods not accounted for, and with very little control of validity and reliability. Nevertheless, such data, for example in the form of ratings and rankings, can play a functional role as a sort of quick orientation for consumers and others. As such they have considerable success as a social form (Esposito and Stark 2019).

Which Power Effects of Evaluative Information are Made Possible by Circulation?

To capture the power effects of evaluation, we need a way to describe how evaluation interferes with social reality not in the form of brute force, but also not just as a result of independent voluntary choice.

There is an increasing literature which shows how evaluative information such as performance measures exerts influence upon public services and practices (Smith 1985; Munro 2004; Ozga et al. 2011, Davis et al. 2012). Dahler-Larsen (2014) proposes a conceptual framework where so-called constitutive effects can be identified in the following domains: the content of work, social relations, timing of activities, and the broader world view which defines the meaning of work. He also argues that these effects might be so deeply socially constructive that it is little fruitful to categorize them as merely “unintended”, as if some clear, definite, and original intentions behind such systems were the only relevant yardstick for assessment. Instead, in complex organizational systems, effects of evaluative information can be pervasive, complex, non-linear, multi-layered and longstanding. Yet, they do not occur randomly. They are produced inside strategic landscapes defined by large social systems, and they are neither a result of brute force nor independent voluntary choices.

Performance measurements create new realities that are morally ambivalent (Power 2016). For example, some see people “gaming” the system, while others see the same professionals using registration practices to find ways forward for their clients and patients which would be otherwise impossible (Kerpershoek, Groenleer and de Bruijn 2016).

The role of managers in relation to a given performance evaluation system can be proactive or reactive. Managers in some universities put more emphasis on particular performance indicators in research evaluation in their own internal university management than legislation required them to do, perhaps because it allowed them to increase the overall scores of their institution and/or their own position as managers (Lind 2019). In this way, the same national evaluation system can have quite varied local effects.

The effects of performance information are often not fully predictable. Some indicators can be functionally used in practice even if most of the involved

people recognize their limitations regarding validity and reliability (Dambrin and Robson 2011). The “liberation” from methodological constraints may in fact only enhance the pragmatic value of evaluative information (Esposito and Stark 2019). In some situations, people seek to influence the purpose, design and use of particular indicators (Pollock et al. 2017) although some are in better positions to do so than others, and although this kind of involvement may be seen as morally ambiguous (Jensen 2011). It can therefore be difficult to decide how much one should engage in the production, design and use of evaluative information.

A lack of clarity around the potential effects of performance indicators is not necessarily the same as a weakening of their effects. Moral ambiguity and causal uncertainty may in fact trigger anxiety in people which makes them prepared to change their strategies and practices.

In other words, the effects of evaluative information and the uncertainty of the social processes in which these effects occur are not necessarily two separate phenomena. People may treat uncertainty of one’s future performance scores as a “risk factor” which needs to be subject to “risk management” (Power 2016). In this light, people take precautionary steps to make themselves and their organizations auditable. For example, a program for students will be planned in a way where (future) student satisfaction scores are “properly managed.”

Thus, a number of socially constructive processes may help turn inscriptions of performance into processes of social construction of reality (Power 2019). While some mechanisms such as self-fulfilling prophecies may help accelerate this process, it should not be forgotten that these social processes are capricious and not automatic (Power 2019). To Power’s points it can be added that social circulation of performance scores is probably a strong conducive factor, as it sets in motion a dynamic interplay between forces which are external and internal to organizations as well as external and internal to people. On this basis, we can articulate a theoretical idea: *To the extent that people foresee how important circulation is for the effects of evaluative information, it is likely that aspects of circulation (infrastructures, rules, and motivations, etc.) will be something that people are interested in, care about, and struggle about. Or in other words: If there is power in circulation, people will seek to exert influence on the process of circulation.*

In the following, two cases are presented which illustrate this theoretical idea in very different ways.

Case 1: Google Scholar

Google Scholar is a feature provided online by Google where the number of citations of a scientific book, article or other product, is counted. In addition, any author of scientific work can relatively easily establish a personal Google Scholar webpage, where the number of citations of all works of that author are shown and counted on one page together with a few metrics (such as number of citations over the last 8 years), H-index and more. By default, publications are listed in the order of their number of citations. By clicking on a given text, it is easy to see a list of the actual citations of an author’s work.

The evaluative information provided by Google Scholar has been criticized on both conceptual and methodological grounds. It is uncertain whether the number of citations measures academic impact, and how impact is defined more specifically. It is also debatable whether impact hinges on the quality of the research, the networks of the particular scholar, the language in which he or she publishes, the circulation of the channels in which he/she publishes and whether it is meaningful to compare citations counts of researchers without normalization of their scientific fields and subfields.

On methodological grounds, Google Scholar has been criticized for lack of transparency in general, including lack of master list and lack of description of quality control (Delgado-Lopez-Cozar and Cabezas-Clavijo 2013: 105; Halevi, Moed and Bar-Ilan 2017: 830).

A person working at Google Scholar argues that the evaluative data in Google Scholar are self-validating to some extent, because if scholars cite scholars, they are presumably citing something of scholarly value (van Noorden 2014).

However, the fundamental problem regarding the lack of transparency regarding the underlying procedures leading to one citation count rather than another remains unsolved. Nevertheless, Google Scholar is used to an increasing extent for number of purposes. Let us therefore unpack some of the drivers behind the circulation of Google Scholar data.

Allegedly, there is no “business model” undergirding Google Scholar itself but it may help promote a positive view on Google itself among “some of the smartest people in the world” (!), the academics, and it is easy and cheap to operate (van Noorden 2014).

A quick search may allow someone to quickly see how cited a given researcher is, which is a key source of academic status, even if there is debate about the meaning of citation counts (as provided by Google).

Whether a researcher decides to make a Google Scholar profile and make it publicly available for others is therefore likely to be influenced by whether this move is believed to enhance one’s reputation. Given that each citation is counted, Google Scholar allows its users to distinguish between the scores of two researchers with great granularity, e.g. 3456 against 3565 citations.

However, as more researchers make such a profile available, it is possible that those who do not will be subject to suspicion that their citation scores are not very good regardless of whether that is true.

A researcher who opens a personal profile in Google Scholar can subscribe to regular updates about the citations he or she receives from other researchers. This may boost one self-confidence, serve as a motivation factor, and also keep one informed about researchers who have similar interests, who might be potential partners in collaboration or whose work should be followed for inspiration. The Google Scholar profile is thus able to serve several purposes at the same time, such as motivation, promotion, and network-building.

In this way, several mechanisms are working together. These mechanisms operate regardless of whether Google Scholar also has an official role to play for example in research evaluation systems in a given university management structure. Of course, it can be a part of a university’s research evaluation policy

to systematically require researchers to establish such profiles and/or to trim them in particular ways. Nevertheless, researchers can make independent decisions about the publication of their individual profiles without any formal rule requiring them to do so. But they do so in the broader context of the increasing importance of competition and self-promotion among researchers. A focus on citation scores may not only have consequences for marketing of one's work but can also have constitutive consequences for the work itself. Using simple math, a researcher can project one's score for the ongoing or coming year and increase strategic efforts accordingly. An advanced analytical mind can also make decisions about future publication strategies based on the number of citations of different publications with different topics in different outlets. One might decide, for example, that publications in one's own language rather than in English have too little effect (Dahler-Larsen 2018). Analysis of Google Scholar citations can thus be used to develop a research strategy aimed to enhance one's Google Scholar citations.

Evaluative information will be circulated in a large scale if each person who is engaged in its production, promotion and use also 'inspires' at least one other person to do the same. There are very few rules which regulate the circulation of Google Scholar. Although the designers at Google Scholar may have done a good job in inventing a smooth set of features, the centre of calculation is not in position to secure global circulation. Google Scholar would not have reached its present level of widespread use were it not for its ability to create an alignment of interests between the evaluation machinery and its many individual users.

The logic may be compared to that of a virus. All that is needed for a virus to spread globally is that each person infected spreads the virus to more than one person ($r > 1$).

If this analysis is correct, it has implications for an analysis of power. Individual users of evaluative information share a little fraction of the power inherent in the wider circulation of this information which, in turn, helps enforce its social influence. These individuals are not isolated, of course. Their actions and decisions are embedded in a larger organizational and socio-political research regime which thrives on both networking and competition.

Case 2: Workplace Assessments

Directive 89/391/EEC (European Agency for Safety and Health at Work, 1989) established a common framework to secure a minimum level of protection from work-related health and safety risks for workers in all member states of the European Union (European Commission 2017). The framework articulates the minimum standards and broad principles for "the prevention of occupational risks, the protection of safety and health, the elimination of risk and accident factors, the informing, consultation, balanced participation in accordance with national laws and/or practices and training of workers and their representatives, as well as general guidelines for the implementation of the said principles" (European Agency for Safety and Health at Work 1989: 4). It covers all workplaces, risks and sectors of activity, both public and private.

The Directive places responsibility for risk assessment and risk management on the employer. In addition to complying with minimum legal standards,

employers must also actively carry out risk assessments and decide on the improvement measures that best meet the identified risk profile of their company (European Commission 2017: 10). Given this wide discretion for employers concerning the design and use of the stipulated risk assessment, the form of this regulation can be characterized as soft (Smismans 2003). The Danish WPA legislation is a direct consequence of Framework Directive 89/391.

Workplaces are legally required to carry out a workplace assessment (WPA) at least once every three years. They must consider workplace health and safety issues and include a documented action plan that is made available to managers, employees and inspection authorities, but it is not made publicly available.

The form and shape of WPAs is discretionary. For organisations with thousands of employees, the form of the WPA may be determined by upper-level management, the HR department, etc. Others delegate much discretion to local actors in the organisation (e.g., local bosses, employees and local work environment representatives (WERs) elected by their colleagues). Some seek the help of consultants or accreditation agencies (Hohnen and Hasle 2011). As a consequence, the WPAs even in the same type of organization can be very diverse. Because of the soft nature of this regulation, WPAs are described in the public imagery as everything from a useful tool to a ritualistic tick-box exercise.

In a recent evaluation of WPAs among 2221 workplaces in the public sector, 52.4% of the WERs found the WPA to be a “very useful” instrument (Dahler-Larsen and Sundby 2019: 108).

However, 13.2% of the WERs said that the most recent WPA did not lead to an action plan. Additional 8.6% said they did not know whether such action plan existed. In sum, 21.8% of the WERs were therefore not able to use an action plan to back up the most recent WPA. Not only is an action plan legally required. An action plan was also found to be a strong predictor of subsequent action taken as a result of the WPA (Dahler-Larsen and Sundby 2019). A lack of an action plan as part of the WPA is therefore a major problem in the implementation of the WPA legislation.

In order to ameliorate this problem, it was suggested to make WPAs publicly available (Dahler-Larsen 2019a). It was argued mandatory publication of WPAs would:

- i. serve to remind the employer to actually carry out WPAs and make action plans
- ii. allow inspectors to easily check whether legislation had been followed
- iii. allow WERs to easily verify whether they had an action plan or not
- iv. increase the commitment to follow-up actions
- v. allow job applicants to get insights into the quality of the work environment in their future job before applying
- vi. allow good examples of workplace interventions to spread more easily
- vii. allow good examples of WPAs to spread more easily
- viii. add pressure of workplaces with a very bad score in a WPA to actually take action
- ix. generally sustain more focus on the importance of a good working environment

The proposal generated debate in meetings and conferences. Managers did not support publication of WPAs. Their main argument was that WPA is an

instrument designed for internal purposes (the assumption apparently being that this evaluative instrument can only be used for the purposes for which it was originally designed in the 1980'ies).

Managers also fear to be blamed and exposed if the WPA scores in their organization is comparatively worse than others. Even some of the trade unions which represent work environment representatives have only lukewarm support for the proposal, perhaps fearing that the WERs themselves will also be left "hanging to dry" in the eyes of the public. Others believed that an offensive policy on WPAs would be the best way to increase focus on work environment issues in general in society.

It was also argued that publication of WPAs would be in conflict with GDPR legislation. The argument does not hold since WPA data are required to be available for all employees in an organization, so GDPR-sensitive data would have to be filtered out anyway. It was also argued that WPAs from different organisations would be difficult to compare because of idiosyncratic nature of the data produced in each organization's self-evaluation of the WPA. In real life, however, many organizations use external consulting companies who use standardized survey instruments to produce WPAs. So, in fact, it is easy to compare the findings across organizations. This is done with many other types of data.

For example, data about student satisfaction in higher education institutions is made publicly available through a ministerial website. The consulting company producing these data is the same as the one who produces WPA data for University of Copenhagen. However, while the first set of data is publicly available (uddannelseskvalitet.ku.dk), the other is only available at an internal University of Copenhagen website. As a consequence, for example, students' evaluation of the psycho-social study environment is widely circulated, while employees' evaluation of their psycho-social work environment is not. This difference between these two regimes cannot be explained by the alleged technical and methodological arguments concerning principles of comparability. Clearly, the two domains of psycho-social environments are not ontologically different objects. In fact, the formats for data presentation are strikingly similar, all the way down to visual design and the colour coding of responses on five-point scales. Yet, the two types of data are placed under two different regulatory regimes regarding distribution and circulation.

The lack of institutional rules requiring mandatory publication of workplace assessments do not prevent other activities with data. For example, consulting companies who help public organizations carry out WPAs maintain ownership of the data after a stylized report is sent to the workplace. (This is one of the ways to assure personalized WPA data do not float inadvertently around at the workplace). As long as the consulting company does not reveal the identity of the workplace or the respondents, nothing prevents that company from using the data in further analyses. Since the same survey questionnaire is often used repeatedly, a consulting firm can build up a database which allows benchmarking across all public organizations with certain background characteristic, such as type of organization or geographical region. In the next step, this feature can be used as a sales point so that future customers can see their own scores against those of comparable organizations. In this way, WPA

data can circulate, although the link between any given organization and its data is never revealed publicly. While it can be argued that the public sector will thereby pay consulting companies two times for data which the public sector in fact delivers in raw form, there is not much public debate about this arrangement, since it is invisible for the public.

In sum, in spite of the massive amounts of data being collected, there is no public circulation of WPA data which allows one to identify workplaces with a very good versus a very bad working environment. The struggles resulting from WPA data all take place within each organization. This fact restricts the extent to which variations in WPA scores can be articulated and addressed as a political issue.

As a consequence of the whole debate about publication of WPAs, a journal conducted an opinion poll about the publication of WPAs. A large majority appeared to be in favour of publication of WPAs (52.6% yes against 22.9% no, while 24.5% did not know what to say).

Analysis and Discussion of the Cases in Terms of Infrastructure, Incentives, and Rules

It was argued earlier that three interrelated aspects are particularly critical in relation to circulation of evaluative information: the infrastructure which makes circulation possible; the interests and incentives connected to various roles such as ‘object’, ‘provider’, and ‘user’ of evaluative information; and the institutional rules which regulate circulation. Let us systematically analyse the two case narratives in terms of these three categories.

Infrastructure

Infrastructure enhances circulation if it produces a uniformity of evaluative traces and inscriptions from the outset. Without such standardization, further circulation becomes not impossible, but bedevilled by frictions.

The inventors of Google Scholar have designed a web-crawling device which identifies ‘a scientific paper’ and ‘a citation’ in ways which are not entirely clear, but works well enough to provide a rough automated estimate of real citations. Given Google’s position as a provider of services on the internet, it has probably been relatively easy to secure worldwide circulation of Google Scholar citation data.

While infrastructure in principle allows WPAs to be easily published on the internet (and most, if not all, modern workplaces have a website), the unstandardized way in which the production of WPAs takes place is used as an argument against circulation. Each organization is allowed to use its own format for design and data collection. Although WPAs could, in principle, still be made public (in some cases law requires evaluations to be made public even if they are not made in compatible formats (Dahler-Larsen and Sundby 2019)), the uneven formats of WPAs are used both explicitly and implicitly as an argument against circulation. The WPA in ‘our form’ is simply perceived as ‘our WPA’.

The whole landscape of infrastructural forces is more complicated than that, however, because some general results from those WPAs which are made with the same format made by consulting companies are in fact circulated for further

commercial use without an explicit link between each public organization and its data. This observation testifies, again, to the importance of infrastructure. In the context of a market for evaluative data, there is an incentive for consultants to further exploit data infrastructures for commercial purposes. Each company who purchases this information will know its WPA score relative to other (anonymized) scores, but will also use it only for internal purposes. If all the WPA information in the country were publicly available, its commercial value would be undermined. The infrastructure controlled by consulting companies influences the way in which data can be circulated and the format of the resulting data, but their use of this particular infrastructure is also made possible, incentivized, and restricted by the broader context of the market.

Interests and Incentives

Why does Google produce Google Scholar data free of cost? While Google itself makes most of its money on advertising, it has been suggested that Google Scholar helps create an image of a positive and useful brand among academics as an attractive group of users. The costs of running Google Scholar may be surprisingly low. While it takes some ingenuity to invent and refine the underlying algorithm, its daily operation is presumably entirely automatic.

The key to circulation lies in the interests of individual users to use Google Scholar to get a quick and easy overview over a researcher or a paper and its citations. The fact that the total number of citations for a researcher is easily available for every researcher with a personal profile (in spite of the researcher's age, research field, and other factors you would take into account in a more considerate form of evaluation) makes it easy to access and digest data at first sight. In turn, individual researchers may find it fruitful to have a personal Google Scholar profile to enhance self-motivation, reputation-building, and networking. Google Scholar rewards researchers with such a profile by sending ongoing information about who has cited one's work. Aside from supporting the researcher's self-esteem, it may also help make fruitful connections with researchers around the world in a way that is faster and more reliable than browsing through journals.

What Google Scholar actually measures remains ill-defined, except that it is "Google Scholar citations" (Delgado-Lopez-Cozar and Cabezas-Clavijo 2013: 105; Halevi, Moed and Bar-Ilan 2017: 830), but the debate about this issue has not stopped the circulation of Google Scholar data. As more people use it and more reputable researchers have their own profile, it becomes more difficult to argue that Google Scholar is without validity at all. Furthermore, once researchers take control of their own profiles, it is likely that they will help clean out some duplicates, faulty registrations, papers not written by themselves, and other unfortunate phenomena which otherwise afflict machine-made internet registrations. There is an individualized interest to improve data.

The situation could not be more different with a WPA. It is legally defined as a collaborative exercise with full partnership from both employers and employees. One partner cannot move far with WPAs without the consent of the other, but their interests are often not overlapping, particularly not in the direction of publication.

Most employers would usually not want to be held accountable for the WPA results. The WERs would probably see it as a strong normative violation of the cooperative spirit intended in the WPA process if they published the results of a WPA without the consent of the employer. They might also fear that some would blame the WER as partly responsible for the work environment. Last, but not least, the WER is usually not in a very powerful position at the workplace. The alliance between an evaluation system and the power of an internal advocate of that system is a key predictor of its social power (Power, 2019). People with higher education are under-represented in the position as WER (Dahler-Larsen and Sundby 2019: 54). Any wider use of the evaluative information is weakened from the outset because WPAs do not have a bridgehead inside the organization in the form of a strong and reputable internal advocate (Power 2019). It is not easy to build alliances with external partners. The working environment is not a highly profiled policy area neither internationally nor nationally (Smismans, 2003).

Rules

The circulation of Google Scholar is not particularly promoted by legal rules or management decisions. Instead, it hinges on a smooth set of infrastructural machineries in combination with a particular alignment of individual interests in promoting and using the resulting evaluative information. All that is needed is that no institutional rule stands in the way of automatic crawling through already available information and circulation of the results. At later stages, university managers and others may “implement” Google Scholar in their own internal rules for academic evaluations and assessments. The “decision” to circulate Google Scholar is taken nowhere and everywhere, in Silicon Valley and on the keyboards of computers around the world, and in the brains of researchers and others. Perhaps Google Scholar points to the future by showing how evaluative information can be circulated while leaving governmental and democratic processes to play a secondary role as implementors, at best.

Again, the situation could not be more different with WPAs. From the outset, WPAs were legally defined as a tool for internal learning and reflection. A configuration of infrastructures, interests and incentives makes sure that WPA data do not circulate, even when it is technically possible to do so.

A long-term consequence of the lack of publication of WPAs is that policies in the area of the working environment continue to suffer from lack of evaluative data (Smismans 2003). The alleged “lack of data” helps sustain a vicious circle in an era where policy change is expected to be justified on the basis of the provision of systematic evaluative data (Radaelli and Meuwese 2010).

Given this constellation of interests and positions, if WPAs are to be made publicly available, the last resort would be a change in general political rules about WPAs. That is why controversy has emerged over exactly this issue. However, the work environment is not a highly prioritized policy area, and change regarding publication of WPA data may not be likely even if a majority of the population supports it.

Conclusion

As evaluation is thus partly emancipated from both representative democratic bonds and methodological foundations, it is particularly relevant to study circulation of evaluative information both as a result of power and as a source of power. Like Espeland and Sauder (2016), we have argued that evaluation may have serious constitutive effects. The power to produce constitutive effects rests to a large extent with the power inherent in circulation of such evaluative data. As a consequence, power is also exercised in attempts to curb circulation. Circulation is best studied as fragile and contested phenomenon. While more attention to this phenomenon is both theoretically and practically important, a departure from some forms of conventional thinking is also needed.

It is no longer enough to focus on the methodological aspects of evaluative data, as evaluation courses and textbooks often do. In our cases, classical methodological criteria such as validity and reliability (and meta-data which permit assessment of validity and reliability) were not the key factors determining whether circulation took place or not (a finding consistent with Esposito and Stark 2019).

Furthermore, studies of evaluation in public administration should not focus only where there is official, policy-based regulation of publication of data. There are also empirical situations where infrastructures of evaluative information align with particular coalitions of interests in a way that makes classical, political and managerial regulation of the circulation of evaluative information obsolete, less influential, or secondary. Research on evaluation should be able to grasp both of these contingencies.

Finally, a broad social criticism of the ideal of transparency as dominant in our society may be relevant up to a point. Sometimes, the ideal that everything should be transparent results in tyranny (Mueller 2018), but the critique of transparency often confuses its alleged object with empirical reality. Some evaluative data are not circulated. The critique of the present culture of transparency misses the situations where evaluative information is not made public, and where the process leading up to this situation is neither transparent nor easy to understand (WPAs) as well as the situations where no meta-data are circulated about published data (Google Scholar). Yet, these “negative” cases, ie. the absence of transparency, may be as significant for politics, for public management, for professional practices, and for citizens, as the “positive” examples. While or more precisely because the constitutive effects of evaluation are powerful, there are also attempts to channel, regulate or even stop certain streams of evaluative information.

Therefore, theoretical studies should treat circulation of evaluative information as a variable and as a contested social process which depend on many factors including those categorized as “infrastructures”, “incentives”, and “rules” in the analysis presented here. Future studies will be interesting to follow. It will also be interesting to see whether it is possible to invent democratic structures which provide feed-back regarding the constitutive power effects of evaluation into the same increasingly complex social relations and structures in which evaluative information is circulated. At the same time, of

course, the very conditions which make circulation possible should also be seen as a part of the problem.

Acknowledgements

This article is a product of a larger project funded by the Danish Working Environment Research Fund.

References

- Bernstein, Ethan (2017) Making Transparency Transparent: The Evolution of Observation in Management Theory, *Academy of Management Annals*, 11 (1): 2017 – 266
- Brown, W. (2017) *Undoing the Demos*, Zone Books, New York.
- Dahler-Larsen, Peter (2012) *The Evaluation Society*, Stanford University Press, Redwood City
- Dahler-Larsen, Peter (2014) Constitutive Effects of Performance Indicators: Getting beyond unintended consequences, *Public Management Review*, 16 (7): 969 – 986
- Dahler-Larsen, Peter (2018) Making citations of publications in languages other than English visible: On the feasibility of a PLOTE-index, *Research Evaluation*, 27 (3): 212 – 221
- Dahler-Larsen, Peter (2019a) Arbejdspladsvurderinger er et lukket land, *Politiken* (11.10.2019), Debat: 6
- Dahler-Larsen, Peter (2019b) *Quality: From Plato to Performance*, Palgrave Macmillian, London
- Dahler-Larsen, Peter & Anna Sundby (2019) *Arbejdspladsvurderinger – hvordan et obligatorisk virkemiddel til evaluering af arbejdsmiljøet virker på offentlige arbejdspladser*, Syddansk Universitetsforlag, Odense
- Dambrin, Claire & Keith Robson (2011) Tracing performance in the pharmaceutical industry: Ambivalence, opacity and the performativity of flawed measures, *Accounting, Organizations and Society*, 36: 428 – 455
- Davis, Kevin E., Angelina Fischer, Benedict Kingsbury & Sally Engle Merry (2012) *Governance by Indicators: Global Power through Quantification and Rankings*, Oxford: Oxford Press
- Delgado-Lopez-Cozar & Álvaro Cabezas-Clavijo (2013) Ranking Journals: Could Google Scholar Metrics be an Alternative to Journal Citation Reports and Scimago Journal Rank?, *Learned Publishing*, 26 (2): 101 – 114
- Espeland, Wendy Nelson & Michael Sauder (2016) *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*, Russell Sage Foundation, New York
- Esposito, Elena & David Stark (2019) What's Observed in a Rating? Rankings as Orientation in the Face of Uncertainty, *Theory Culture and Society*, 36 (4): 3 – 26
- European Agency for Safety and Health at Work (1989) Directive 89/391/EEC – OSH “Framework Directive”
- European Commission (2017) *Ex-post evaluation of the European Union occupational safety and health Directives*, European Commission, Brussels

- Feldman, Martha S. & James G. March (1981) Information in Organizations as Signal and Symbol, *Administrative Science Quarterly*, 26 (2): 171 – 186
- Fochler, Maximilliam & Sarah de Rijcke (2017) Implicated in the Indicator Game? An Experimental Debate, *Engaging Science, Technology, and Society*, 3: 21 – 40
- Halevi, Gali, Henk Moed & Judit Bar-Ilan (2017) Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation – Review of the Literature, *Journal of Informetrics*, 11 (3): 823 – 834
- Hesselmann, Felicitas & Cornelia Schendzielorz (2019) Evaluations as value-measurement links: Exploring metrics and meanings in science, *Social Science Information*, 58 (2): 282 – 300
- Hohnen, Pernille & Peter Hasle (2011) Making work environment auditable: A ‘critical case’ study of certified occupational health and safety management systems in Denmark, *Safety Science*, 49 (7): 1022 – 1029
- Jensen, Casper Bruun (2011) Making Lists, Enlisting Scientists: The Bibliometric Indicator, Uncertainty and Emergent Agency, *Science Studies*, 24 (2): 64 – 84
- Kerpershoek, Emiel, Martijn Groenleer & Hans de Bruijn (2016) Unintended Responses to Performance Management in Dutch Hospital Care: Bringing together the managerial and professional perspectives, *Public Management Review*, 18 (3): 417 – 436
- Kroll, Alexander (2015) Drivers of performance information use: Systematic literature review and directions for future research, *Public Performance and Management Review*, 38 (3): 459 – 486
- Latour, Bruno (1987) *Science in Action: How to follow scientists and engineers through society*, Harvard University Press, Cambridge, MA
- Latour, Bruno (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory*, Oxford University Press, Oxford
- Law, John (2004) *After Method*. London: Routledge.
- Ledermann, Simone (2012) Exploring the necessary conditions for evaluation use in program change, *American Journal of Evaluation*, 33 (2), 159 – 178
- Leeuw, Frans & Jan-Eric Furubo (2008) Evaluation Systems What Are They and Why Study Them? *Evaluation*, 14: 157 – 169
- Lind, Jonas Krog (2019). The missing link: How university managers mediate the impact of a performance-based research funding system, *Research Evaluation*, 28 (1). 84 – 93
- Mau, Steffen (2019) *The Metric Society: On the Quantification of the Social*, Polity Press, Cambridge
- Merry, Sally Engle, Kevin E. Davis & Benedict Kingsbury (eds) (2015) *The Quiet Power of Indicators: Measuring Governance, Corruption and Rule of Law*. Cambridge: Cambridge University Press
- Mueller, Jerry Z. (2018) *The Tyranny of Metrics*, Princeton University Press, Princeton NJ
- Munro, Eileen (2004) The Impact of Audit on Social Work Practice, *The British Journal of Social Work*, 34, (8): 1075 – 1095

- Osterloh, Margit & Bruno S. Frey (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
- Ozga, Jenny, Peter Dahler-Larsen, Christina Segerholm and Hannu Simola (eds) (2011) *Fabricating Quality in Education: Data and Governance in Europe*, Routledge, London
- Patton, M.Q. (1978) *Utilization-focused evaluation*. Sage, Beverly Hills.
- Peters, Brainard Guy & Jon Pierre (2020) From evaluation to auditing and from programs to institutions: Causes and consequences of the decline of the program approach, *Governance*, 33 (3): 585 – 597
- Pollock, Neil, Luciana D'Adderio, Robin Williams & Ludovic Leforestier (2017) Conforming or transforming? How organizations respond to multiple rankings, *Accounting, Organizations and Society*, 64: 55 – 68
- Power, Michael (1997) *The Audit Society: Rituals of Verification*, Oxford University Press, Oxford
- Power, Michael (2019) Modelling the Microfoundations of the Audit Society: Organizations and the Logic of the Audit Trail, *Academy of Management Review*, 0, DOI: <https://doi.org/10.5465/amr.2017.0212>
- Power, Michael (ed) (2016) *Riskwork: Essays on the Organizational Life of Risk Management*, Oxford University Press, Oxford
- Radaelli, Claudio M. & Anne Meuwese (2010) Hard Questions, Hard Solutions: Proceduralisation through Impact Assessment in the EU, *West European Politics*, 33 (1): 136 – 153
- Rist, Ray C. & Nicoletta Stame (eds) (2006) *From Studies to Streams: Managing Evaluative Systems*, Transaction, Brunswick, NJ
- Roberts, John (2017) Managing *only with* transparency: The strategic functions of ignorance, *Critical Perspectives on Accounting*, 55: 53 – 60
- Rosa, Hartmut (2014) *Fremmedgørelse og acceleration*, Hans Reitzels Forlag, Copenhagen
- Smismans, Stijn (2003) Towards a new community strategy on health and safety at work? Caught in the institutional web of soft procedures, *The International Journal of Comparative Labour Law and Industrial Relations*, 19 (1): 55 – 84
- Smith, Peter (1995) On the unintended consequences of publishing performance data in the public sector, *International Journal of Public Administration*, 18 (2-3): 277 – 310. doi: 10.1080/01900699508525011
- Thomsen, Jens Peter Frølund (2002) *Magt og indflydelse*, Aarhus Universitetsforlag, Aarhus
- Van Noorden, Richard (2014) Google Scholar Pioneer Reflects on the Academic Search Engine's Future, *Scientific American* (11.10.2014)
- Vedung, Evert (1997) *Public Policy and Program Evaluation*, Transaction Publishers, Piscataway, NJ
- Weiss, Carol H. (1977) Research for Policy's Sake: The Enlightenment Function of Social Research, *Policy Analysis*, 3 (4): 531 – 545