

Rasputin's black box

How statistics are used in Norwegian municipality monitoring and administration ¹

Norwegian municipality politics- and administration is experiencing a downscaling from hierarchies towards two-level-models. With up to 250 divisions on the second level, communication of expectations and reports are simplified through statistical tools and expressions. How is this quantification of monitoring affecting the ways of government?

The basic assumption is that statistics and methods are historical, social and most of all political constructs, results of conflicts and compromises through the last 200 years. What is quality, how do we monitor it and describe it? Description through numbers - and which way to describe through numbers - are choices of political nature. We believe the consequences of Norwegian municipalities' use of statistics fall into two parallel categories: Changes of ways of seeing political issues, and increased distance in decision-making.

The study takes basis in qualitative data of two administrations, and a broad survey of all Norwegian municipality administrations. The part of the study presented here discusses the tools and methods in use, and through discourse- and rational choice analysis we discover how these technologies both change and manifest ideas, values and ways of government. We use the concepts of standardization, quantification, statistical analysis and formalization to specify the idea of statistics as an analytical object, and find the empirical concepts of monitoring and audit as describing parallel processes through which an audit culture is brought forward in the organizations.

Jon Hovland

Department of Sociology and Political Science

Norwegian University of Technology and Science NTNU

N-7491 TRONDHEIM

Norway

jonh@svt.ntnu.no

1 Introduction

Audit cultures has since the last fifteen years made their way into managements and public policies, and through the last ten years also into the writings of social scientists (Power 1997; Strathern 2000a). This paper reports how some of the way into the public administrations arises through a distortion between matters of monitoring, that is internally validated, and audit, that is externally validated. The exact meaning of these two concepts will be thoroughly

¹ This paper is a part of the research program "For whom the Bell Curves" (<http://www.svt.ntnu.no/iss/projects/bell>), financed by the Norwegian research Council, and coordinated by Prof. Ann Rudinow Sætнан at the Norwegian University of Science and Technology, Trondheim. I have also had great help from supervisor Svein Hammer, and colleague Nadji Aïssa Khefif.

explained through the analysis, it is a matter of standardized evaluations, and how one finds these useful.

To clarify the important distinctions between these two concepts, the paper gives a thorough introduction to an analysis tool to understand processes concerning statistics that theoretically detaches standardization, quantification, statistical analysis and formalization. This is both to explicate that the focal point of the study is statistics, and then again what statistics mean, and also to use these concepts, or elements, to illuminate the meaning and consequence of the difference between audit and monitoring. Through this, we find in this paper that a dialogue concerning monitoring and arguments of standardization, may transmit to audit and arguments of quantification. This way the administration gains a technology of governing through audit that builds on a consensus, not for audit, but for monitoring.

These findings are based on studies of Norwegian municipality administrations that have experienced downscaling and reorganizing of the organization (a general and strong tendency in contemporary Norwegian administrations), and through removing mid-levels of authority observe an enlarged gap between strategic top-level authority, and operative unit-leaders². It has been the suspicion of my project group that this gap is filled through dialogue routines and statistical monitoring in a way that serves as an even more powerful and robust technology of governing. This paper does not provide answers as to the strength and robustness of this way of doing government, but does suggest processes through which the prompting of such technologies is possible.

The first part provides an introduction to the way of understanding and analysing what I term ‘The ghost of statistics’, followed by an introduction to Norwegian municipalities and the data used in this study, and thereafter an analysis where *audit* and *monitoring* is used to show differences in understanding statistics in the organization (I term them languages), and these two are associated with typical arguments of two out of the four ways of analysing statistics as described in the first section. In discussing the concepts, or elements of statistics we will bracket the idea of monitoring vs. audit, only to return to them in the analysis. For now, we will focus on general ideas of statistics

² To be exact, the gap itself is not widened, but the range of units controlled by the higher level is greatly expanded. For all practical reasons it would be likely to assume that the gap thereby also is widened.

2 The ghost of statistics

Studying the use of statistics in public affairs is a tricky task. The reason why becomes clear both in inquiring the field and the literature. When confronting informants in the organisations, we'd be met first by a round of clearing up: "Ok, what kind of use of statistics do you mean?" Referring to specific statistics we know are used, this problem is overcome. However, interrogating literature the question re-emerges. "How is this useful for me? Is this article actually referring to statistics, or, what is it?" Again it moves down to the analytical level: "What do I study? What do I discuss with which sets of arguments?" My conclusion to all these questions was that we were addressing a ghost, something that disappears into seemingly nothing as soon as you get it into eye-sight.

This is as much a substantial problem as a theoretical problem. Substantial, in that this "nothingness" of statistics is how they become both as powerful and as friendly-looking as they do, as I will show through this paper, using the more practice oriented concepts of *audit* and *monitoring*. It is also a theoretical problem, in that it is difficult to analyse something that is not identified. To solve the latter problem, I propose a theoretical splitting of the concept of statistics in public affairs into four concepts³: *Formalization*, *Standardization*, *Quantification* and *Statistical Analysis*. These concepts are inspired by Lampland and Leigh-Star's *Formalizing practices*. It should be noted, however, that their project is not to analyze statistics per se, as I do, so I put slightly different meaning into the words, and also add a fourth concept of *Statistical analysis*.

"We have crafted the term "formalizing practices" to serve as an umbrella for grouping together three analytically distinct, but intimately related phenomena: standardization, quantification, and formal representation. Standardizing refers to the broad range of practices—in the economy, science, medicine, politics, and even culture—the purpose of which is to mandate/create comparable and commensurate entities and processes. . . . Quantification, a companion process to standardization, has been accorded more attention by social scientists, historians and philosophers (Hacking, 1990; Mirowski, 1991, 1994; Poovey, 1998; Porter, 1986, 1995; Wise, 1995). This literature has been helpful to build upon in our analysis of standardizing projects. Yet the full analysis of regimes of quantification, their imbrication with all forms of measurement and the daily "doing" of modernity is far from complete. Formal representation needs in turn to be distinguished from quantification, although numbers and mathematics are significant means of formally representing processes and procedures. Numbers, however, do not exhaust the range of formal representations.

³ In other words, if there should be any doubt, this is not a method paragraph; the forthcoming is a description of how one might analyze statistics.

Narrative descriptions of democratic states, used to critique regimes across the globe, are just as much formal representations as are complex ecological models of climate change, though they may differ in conciseness or accessibility" (Lampland and Leigh Star 2006: 4-6, forthcoming, please do not quote)⁴

My project is not to identify historical processes, but rather to cut the different parts loose from each other to make them observable in a narrow context. Therefore I will explain in the following just what I put into these concepts, and why and how I add the fourth concept of *Statistical Analysis*. It should be clear, however, that my inspiration clearly comes from Lampland and Leigh-Star, so that I cannot take full reward for this invention.

2.1 The four musketeers

Before I start describing these four concepts, it should be clear that these are theoretical. They are not meant as real entities, as they are non-detachable to each other. In short, it does not make sense to study them separately, but it is difficult (impossible?) to describe them simultaneously. The first, because they are part of the same logic and the same development; the latter because they involve different arguments, different reasons and effects, different histories, and I would argue different elements of what statistics mean. In the analysis present here I omit their non-detachableness, to return to this matter briefly in its conclusion, and hopefully more explicitly in forthcoming papers. I also point out that the typical debates concerning each concept are not full, that is, there is no extensive discussion of the debates, they are only meant to be indicating so that the reader may recognize the debates from other contexts, and to illuminate that these four concepts have their own lines of argument and their own epistemological histories.

Furthermore I wish to make clear that the succession of these four concepts I propose is of little interest. What came first, and what came after, will in a historical study be a matter of perspective, and in a contemporary study of power structures, as this, they will be observed as parallel processes, driven forward by something that I have not (yet) come to identify. Their co-relationship is unclear to me, and this is probably why I choose not to involve a parallel analysis of the four at this point. To be honest, I wouldn't know what to make of it.

⁴ Please do not quote this quote. It is not yet authorized with the authors. I still chose to include it as I found it far more unethical to pretend the idea to be mine.

2.1.1

Standardization refers to the process through which things are made measurable and comparable. We all know this process very well, through standardization it is possible to know what to expect and how to interpret statements, the obvious example would be SI-standard measures on length, weight, energy etc. Standards turn first-hand knowledge and opinion into visible units, in which anyone with access but limited knowledge can look into the process. Standards also force things into units that may not properly describe them, as when we answer surveys on whether or not we agree to certain statements on a scale from one to 10. Standardization may also refer to the standardization of processes into procedures, such as scripts for telemarketers.

This is probably the easiest accessible criticism when “qualitativists” face “quantitativists”⁵. “Do you really know what you measure? There are different sides of the story hidden in these figures!” These questions very often address the element of standardization in what lies beforehand. In matters of public policies standardization is also abundant. We may find it in regulation of minimum outdoors area for kindergartens, methods of sorting out the most qualified students for studies by middle value grades, and production in service units (like a social security office) by numbers of clients or other selected measures.

The history of standardization is described both by Porter (1995) and Gigerenzer (2001). It’s origins and fundamental reason is opening up the boundaries of inside information, with the ever running risk of throwing out tacit knowledge in the same water. Porter describes how relationship to powerful outsiders and reduced public trust would accelerate processes of standardization, especially through numbers. When the representative of an organization of knowledge and information met the outside world with his information, trust, power and relatedness would be instructive to which degree of standardization he would need in his description.

2.1.2

Quantification refers to the process in which numbers are piled up to columns and graphs. This is obviously dependent on a process of standardization. Still we need to observe this as a separate phenomenon, as the process of standardization may be problematic or may be

⁵ By these notions I mean respectively people who are in favour qualitative methods of research vs. those in favour of quantitative research. This is a rather simplifying distinction and will not be taken any further here.

straightforward. Therefore I will isolate them theoretically, to observe the virtues and problems of the quantification per se.

Hacking and Desrosières describe the process through which quantification became norm through slightly different perspectives. Hacking describes how human nature turned into “normal people with laws of dispersion” (Hacking 1990: vii). From being secret tools of the government, large numbers became public, and in so being turned into statistical laws. Desrosières argues that this development was not a result of scientific development and rationalization, rather result of a pressure from politicians to produce these kinds of numbers to create a level ground for social engineering (Desrosières 1998).

When questioning accounting practices through the concept of quantification, central debates will for instance be on the meaning of the normal distribution: What does it mean that some responses are in the main field, and some are in the outskirts. Does deviance mean non-normal, or non-mediocre, or simply ‘belonging to another scale’? Furthermore, it would concern questions of simple descriptives such as mean values, and what the meaning of middle values are. Is it appropriate in each setting to focus on mean values, or the dispersion for that matter? What meaning do we apply to the units we count and measure by using such terms of evaluation and inference?

Both Hacking and Desrosières also put into their historical accounts how methods of statistical inquiry and inference became the vessels through which descriptive numbers became legitimate ways of generalization and prediction. Desrosières explains how the law of large numbers, the foundation of most of our statistical methods, is a historical mischief: “Its hybrid nature is summed up in the famous and revealing quip dating back to the nineteenth century: astronomers and physicists believe the law to be a theorem demonstrated by the mathematicians, whereas the mathematicians think the law has been proved by the results of repeated testing. In fact, the possibility of conducting multiple and mutually independent observations of comparable objects is the foundation of a statistical methodology initially developed to study populations of human individuals or households.”⁶ (Desrosières 2001: 341).

⁶ This is at least to the specialist no longer a secret, and mathematicians will rightfully claim that probability is mathematics, statistics is not mathematics (Cooke 2005).

2.1.3

However, this leads us on to the elements of *Statistical analysis*, by which I mean how we use quantitative data to infer generalizations and extrapolations. Again, this is meaningless to understand fully without reference to standardization and quantification, and again, the problems baked into it are of a different nature. These are issues of modeling, and how to think about scientific inquiry.

When we are talking about statistical analysis, and maybe criticizing it, we rarely really talk about the analysis itself, as it is a very brief moment in any report, analysis or article, we may term it the statistical moment. We will be criticizing the standardizations, the normal distribution idea (quantification) or how it is used, as we will come back to, but from my experience we will very rarely criticize the actual calculation of inference that is carried out.

If we emphasize the logic of the inference through modeling, we apply arguments concerning both what kind of results models produce, and what kind of process models induce. Statistical models are based on residual testing, that is the value separating each units score in the dataset from the value it would be expected to have from the model or formula we produce. This is in other words a model-internal way of testing, not a testing towards the external world. One could also argue that model working induce a research process in which premises are laid before the analysis is carried out. In other words a process where we specifically define our expectations or hypothesis before we inquire their effects. On the one hand, this produces a straight forward and transparent process, on the other hand it may serve to reinforce categories, as a model-internal way of validating findings will tend to emphasize the meaning of the categories used in the model. Since this kind of argument is not relevant to the meanings of statistics in this study (I will, however, return to this subject in forthcoming papers), we will not enter this discussion any more thoroughly.

2.1.4

All these three seem obviously connected to statistics and society. However, we must be sure to distinguish the two contexts of scientific argument: In society, and inside science (Desrosières 1997). In society, statistics is not only about the method in use, as the three points above point to, but how it is taken further, namely through *Formalization*. When I simplify the original concept of *formal representation* (Leigh-Star and Lampland) to mere

Formalization, it is to rather illuminate formalizing processes in general that are connected to the meaning of statistics.

These are processes through which relations and structures in the organization, such as municipality administrations, are made formally reducible to agreed routines, tests and actions. In one sense, this has nothing to do with statistics, as it is not a matter of calculation, on the other hand it is all about statistics, as the formalizing is dependent on standardizing, generalization and trust in methods, and vice versa: The power of the numbers are dependent on formalizing of relationships. One could argue that just as much as formalization is an element of statistics, statistics is an element of formalization. From that perspective, the findings in this paper could have been interpreted as different forms of formalization. This is, however, a choice of perspective, and for now, and for the present purposes, I choose to use statistics as the focal point of the perspective, and through this formalization could be a concept that should be analyzed as a separate element in the processes observed.

2.2 An analytical tool

This is one out of many analytical methods, according to Porter one could probably do the same study through analyzing the relationship between powerful outsiders, internal integration in the knowledge organization and public trust (Porter 1995). According to Desrosières it would be likely to separate the processes more tightly connected to developments of the state, namely the administrative act, the statistical act and the decentralized act (Desrosières 1997). I am, however, not entirely happy with either of these, as I need an analytical tool not especially applied to explain historical processes, but to analyze contemporary processes more or less taken out of time.

These four concepts are to be regarded as analytical tool-boxes, and when entering the field of public organizations and their use of statistics, the first step will be to identify which box to use. From that point on, one will bracket the surrounding effects connected to the other boxes, to return to them in the discussion. There are obvious problems in such a method, very much connected to the degree to which one uses the magnifying glass (not the mirror, telescope or radar screen) and omit relevant issues. I argue that this is suitable for the purpose in this context, as we here observe closely connected phenomena that each has its own central arguments and positions.

3 Norwegian Municipalities

Norwegian national affairs are organized through 19 county municipalities, that are again split into 435 municipalities. The municipalities's tasks are related to schools, kindergartens, public services, cultural facilities, nursing homes and social security. The municipalities have a parallel organization of a political system, with a locally elected municipal council, and a municipal administration that takes care of daily management under the informed approval of the municipal council. The head of the administration, the city manager is also responsible for strategic organization of the municipality. This study attends the administration, and how it uses statistics in its operational and strategic tasks.

The municipalities have huge differences. They vary in size from 6 to 9700km², and numbers of inhabitants vary from 200 (Utsira) to 540 000 (Oslo). In approximate values, the mean number of inhabitants is 10 000, the median 4500, and mid 50 % of the municipalities have 2500-10 000 inhabitants. Not only do they vary in size, their very different geography and demography makes them entirely impossible to compare. Still there is reason to study them as a whole, given that they are subject to the same regulations and demands from government and law, and they are included in cross-national networks to improve their management.

The last ten years has made the questions of regulations and management improval even more relevant. The Government has since 1997 registered municipal activity through numbers in a large survey (KOSTRA – Municipality-to-Government-Reporting), and introduced Efficiency-networks in which representatives of different administrations meet to compare their numbers and discuss why some are efficient, some are not, according to the numbers. Also the Ministry of Municipality affairs has calculated an efficiency index, in which countable indexes for production for each service is added up to a sum, and compared with the municipal income or funding. By drawing a regression line on production and income, they calculate the degree of efficiency / inefficiency of each municipality from the deviance to the line of best fit.

In addition to this flourishing of calculative practices, there is observably a pressure for reorganization of the administrations, a pressure that probably might be labeled New Public Management-oriented. There have been seminars on Balanced Scorecard (in which you give values to any production and reward thereafter) outsourcing of services, creation of quasi-

markets inside the municipal organization, and in the bulletin for municipality affairs (Kommunal Rapport) there has been a running debate for the last nine years on how to measure quality, with reference to these new calculative practices that have been introduced. There is also a strong demand for and validation of down-scaling of (middle) management and of “liberating” operative units to make their own decisions as to how to meet service production targets while keeping within budget frameworks. According to Norwegian institute for City and Region research (Høvik 2004), this is also seen in a slow but steady increase in number of municipalities that create quasi-markets by enacting a separation between “customers” (the head administration) and “producers” (the units) within the municipal organization, and that have induced “management by objectives” throughout the organization.

These two tendencies, more calculative practices on the one hand, and downscaling through management by objectives, are probably interconnected. What we can say for sure is that they appeared at the same time, their relationship, however, is fuzzy. On the other hand, with reference to the four categories presented in the theory it is easy to see that there are many elements of them all around in this context.

3.1 Data and method

Initially we would use a broad survey of all Norwegian municipalities, and an in-depth study of two-three strategically chosen administrations. However, the results from the survey non-respondents turned too big and also too biased towards certain geographical areas than we could accept for standard statistical generalizations. Therefore the results of this study used with great caution, and only for some descriptive purposes. The main body of findings presented here therefore rests on in-depth studies of two mid-sized city municipalities, both with rather large administrations, in which one has gone rather far in the process of restructuring towards separating between strategic level (top-level) of the administration and operative level (units such as schools, culture facilities etc.) and the other one is in its initial faze of such restructuring. Both are cities with large enough amount of units to make it impossible for the strategic level to have direct control of all unit activities. This gap in operative control is characterized by informants on all levels as a “communicative vacuum”, where midlevel leaders use to be.

The administrations were studied through contacting strategic leaders and gaining access to certain persons involved in statistics and organizational structure, and through these persons finding more employees, networks and groups involved in statistical reporting and work on these reports, and data were collected through interviews (13), one group interview, attending meetings and review of reports and minutes, and documents from other researchers and from the Ministry. Although taking a broad basis, we chose a special focus on primary schools and communication between headmasters and strategic level, as this is expected to be one field with many units, and also much disagreement over reporting and routines. However we did also enter the field of cultural facilities and health services, as these would be the fields of the least and the most standardized reporting routines respectively.

The cities chosen are characterized as “mid-size” in common language, which would mean rather large compared to the calculative average. We found this necessary to allow us to see the effects more clearly, and also rather unproblematic as these cities set the standard for smaller municipalities. In the following text these two municipalities will be merged into one, both for anonymity and for narrative purposes. This is not a study of municipalities in themselves, but rather of certain processes, and therefore this kind of generalization serves the purpose.

4 The two languages

When concerning the school sector, numerous and varied forms of standardized data are collected regularly. Of the most significant, there is the school-home survey, inquiring relationships between the school and homes; there is the pupil-survey, inquiring the pupils appreciation of their school; there is the GSI account of number of pupils, hours of teaching, number of teachers, transport expenses and special needs-pupils; and finally and importantly: The economy accounts. The results of these tests are to be presented to the headmasters regularly through dialogue meetings between the headmaster and a representative of the municipal strategic level authorities. In these conversations level of achievement will be discussed, reasons of failure and success (or reasons why the headmaster finds the numbers irrelevant), and they will discuss the schools eventual achievement of goals for the past year, and goals for the year to come. These goals are to be coordinated with general goals the municipality in dialogue with other unit leaders has drawn out for schools, and for the city in

general. However, the headmaster is responsible for how to attend to, and possibly achieve these goals, and what to make of them.

When referring to these conversations, all sides usually emphasize that the numbers are not important.

“No, they [numbers] tell you nothing, you know” Headmaster

“Well, uh, most of these things [that are to be evaluated] are impossible to measure. [...] It is more a matter of opining.” Strategic level authority.

This is obviously true, dialogue-routines are prevalent throughout the organization to create channels for communication and understanding. But why then all these surveys, and why are they so important to the administrative authority? The answer would only come when we either opened up for explaining possible negative views of statistics, or asked more directly and pushy:

Authority representative: “[...]and then the headmaster must go through all of it [survey reports] and tell me as authority representative, when we’re in that setting [dialogue meeting] ”We also see from this at here, on another subject, there are some things the school has to attend to”. So me, I bring up some major highlights, while the headmaster brings up the rest. I got it all in front of me on the desk, you know”

[...]

Interviewer: ”Do you then relate to mean values and compare to other schools?”

Authority rep.: “Yes, [...]that’s the drawback of it. But what we try to do is, well if you draw this line for the mean value, this is a system with some name, and then we look at how many under and above, and this is what the politicians are interested in.”

As we see, the dialogue is the basic means of communication here, and the headmaster, has good opportunity to state his or her meanings about the statements about the school, as the first quotes above showed, numbers are not seen as very important. On the other, we see that they are the starting point, and it is what the headmaster will have to “answer to”. I will argue that this inconsistency concerning countings that count and countings that don’t count seems to be connected to the double understanding present in the organization on how to understand statistics, and what they are good for. To do this I will introduce the concepts of *monitoring* and *audit*, to see how these may describe different languages through which to understand statistics, see how they are connected to *standardization* and *quantification* respectively, and this may create barriers for dialogue.

Audit is by now a well-established concept in social science. The word has had a striking increase in use since the late eighties, but a definition is difficult (Power 1997: 3-5). It is a matter of systematic cross-evaluation, originated in accounting practices. Even more than it is difficult to define its shape, it is difficult to criticise:

"Yet as an instrument of accountability, holding out the possibilities of a globalizing professional consensus, audit is almost impossible to criticize in principle - after all, it advances values that academics generally hold dear, such as responsibility, openness about outcomes and widening of access."(Strathern 2000b: 3)

Instead of defining audit in itself, I suggest the dichotomy of audit and monitoring. Whereas *monitoring* is the activity in which you survey activities and units to get overview and information, out of the general purpose of evaluating, *auditing* is just the same, but done with a reference to norm or desired value. In other words, monitoring is internally validated, auditing is externally validated.

This is what one may term a "thin" dichotomy; at first glance they may seem to be exactly the same. Let me use an oversimplifying example to explicate the difference. Let's say that I send a draft of this paper for a review with my project coordinator. She would then return it to me with a number of questions and highlightings in the text, and some additional comments and suggestions at the end. Let's take it further and imagine that she does this for a great number of people on a regular basis (which she does), and to simplify her work, she created a system of comments, a number in the text that would indicate what is the problem (which she does not): 1: typo. 2: Etymological problem, meanings of the words do not fit the context. 3: Faulty argument. 4: Important issues missing. 5: Inconsistent structure. 6: Not clear what the paragraph has to do with the argument, and so on. We can also imagine that she on the basis of this sums up what kinds of problems that characterize the text and should be worked on, and uses a standard phrase to indicate where in the process the paper is. This kind of standardization of feedback seems immediately a little awkward, but there is no doubt that I also could use this kind of feedback constructively in my paper. This is paper-internal validating, and I would term it monitoring.

Now let's imagine further that she keeps track of all her feedbacks in a spreadsheet, and uses this to evaluate the state of each of her employees and their progress, and in general feedback explains to each of us how our progress is doing, and asks what we have to say to this. Let's say she uses these dialogues as a basis for reports to the institute and to decide salaries. This is

where monitoring turns into audit, because there is an external validation, and as we seem, the consequences, and the arguments to employ in favor of or against such practice are quite different. Now we will return to the empirical data.

These to very similar, but in effect very different activities (monitoring and audit) are in effect different ways of understanding what is the fundamental role of accounting practices in the municipality administration. With the authority, or strategic level staff, they are both operative:

“[We are trying to] make an assembled report of all evaluations for the politicians, to draw a picture of the [city]-school, the way surveys and statistics show it. [mentions the kinds of surveys]...and then make an assembled picture, of the situation here, over time, and compared to the rest of the nation.” Strategic level officer

This is an example of audit-thinking, where the surveys are not only evaluations in themselves, but measured against external validation. Monitoring, the way I use the term, is also regarded as a central aim for the use of the surveys:

“All headmasters have access to [survey results for their own school]. We try to initiate a local work on each school concerning improvements, and to instigate which groups should join this work.” Strategic level officer

Not only for the units own gain, but also for the authority level, the monitoring qualities of the surveys are emphasized (although the following quote may well be interpreted as lying in the gray area between audit and monitoring):

“...the dialogues concerning [goals and results], and the documentation on how the units do their things [...] has made it easier to take some discussions on leadership. [...]Some unit-leaders may have strived for a long time, and through [this] they get a foundation for reflecting on their own leadership, and maybe find out that they wish to do something else.” Strategic level officer

These are the voices of the strategic level. In the units, represented by the headmasters, the language of monitoring seems to fit well into an established consensus on what school is about. They observe it as a matter of disagreeable pressure from simply being seen and exposed, but on the other hand they find it a right thing to do, as it shouldn't be a secret how relations are at a public school.

“...well that is why the surveys are there, so that we can learn and find out. And then the teachers brought the results into dialogues with every single child and parent. It was very

useful. Very smart. And this is what you really can't be good enough at [...], being more specific in relation to the child, where it stands." Headmaster

"Even though I've been spared for it before, that was not necessarily a good thing for the children at my school. It part of my job to put up with being measured." Headmaster

As we see, monitoring is brought directly into the schools daily routine. Another headmaster reported how the results of school/home-surveys were used as starting points for public debate with parents on how he wanted the school to be, and to gain useful feedback from the parents on his visions and suggestions.

The headmasters feel at home with monitoring, they see it as a natural part of what they do. Although systematic and standardized monitoring increases the pressure, at least the headmasters argue that this is a kind of pressure that they are obliged to tolerate. And when it comes to standardization of the surveys, they find themselves somewhat free to negotiate with the strategic level what to put into the results, and emphasize that they are almost absolutely free to decide what it means for their individual school. It is a tool for them to use.

The language of auditing, however, where the monitoring turns into measures against external standards and norms, does not harmonize with their ideas of what the school should be about.

"...the healthiest kind of measuring [...] lies in where people meet [...] in relevant contexts. [...] Those things that turn up on paper and turn into numbers and parameters becomes ...I think they have some kind of different effect.

[...] They do something to us, these objective - seeming like objective - truths that maybe not really are, influence our thinking. In good ways, making us more on the offensive [...] On the other hand you start thinking in ways like who's fault is this?"

There seems to be a gap here. Whilst the strategic level prompt both monitoring and audit, the operational level (the headmasters) accept only that which is about monitoring. This is also how they communicate statistics in the dialogue with the strategic level authorities. The rest, the audit, is in their eyes for others; for the public, the media and for the administration. It is not their domain, and not something they involve in because they don't seem to believe in it. Although they talk about it, they relate to it as something they do not like. It is not part of the constructive dialogue for creating consensus.

4.2 Two driveways. One gate. One house?

These two ways of understanding and talking about statistics, interface with standardization and quantification as described earlier. It is not related to statistical analysis, and at least not directly to formalization, although that would also be relevant to inquire⁷. When constructing dialogues based on a language of monitoring, it is the essential debates of standardization that will be the foundations for evaluating the system itself. These are matters of whether the measures are correct, whether the survey standards produce priority areas, or in such a case if they produce the *desired* priority areas. If constructing a dialogue based on the language of audit it would rather be the essential debates of quantification that would have predominated. That would be questions of whether the normal distribution is an intuitive way of defining schools, if mean values create irrelevant norms or fortify pressures towards a conformity that may be disadvantageous. Standardization does not equal monitoring, and quantification does not equal auditing, but they seem to be described through these sets of arguments respectively.

It seems that the latter, a dialogue based on audit, is not incorporated in the way these dialogue meetings are carried out, this is not communicated. Both sides refer to this, the headmasters as something undesirable, the strategic level officers as something they do and try to do better. Also, both sides are perfectly aware that these are two different entities. However, it seems not surprisingly to be the situation that the headmasters separate them more clearly than does the strategic part of the administration. Still, although there are two different languages here, that of audit and that of monitoring, there is not a language for saying that these are idealistically different functions or parts of the idea of statistics and surveys. It is my opinion that this is why audit culture can develop the way it does in these administrations, because there is a dialogue concerning matters of monitoring and standardization, but that the consensus produced from this dialogue is transmitted to an area lacking dialogue, namely that of audit and quantification.

This conclusion presupposes a number of conditions, and omits a number of questions: (1) On the analytical level I suppose that the four “elements” of statistics are actually different, and

⁷ As mentioned earlier, one could also analyse formalization, with statistics as one analytical element of that. In such a perspective these would be likely to describe as to different kinds of formalization in the following sense: Monitoring: “Thou shalt talk about this.” Audit: “Thou shalt meet this standard.” The perspective being on statistics as the foundational object of study, I will omit this kind of argument for the sake of clarity.

that they process different logics and arguments. (2) I also suppose that the difference between audit and monitoring is such that it is reasonable to identify either of them. (3) I omit the possibility that the headmasters' acceptance of monitoring is because this in turn can be used as audit over their own subordinates. (4) Also I do not question the "familiarity" with monitoring in the school. Who is it "natural" to monitor, and who has the legitimate power to monitor whom?

5 Rasputin's black box⁸

How does an audit culture pervade municipal administration? I argue that we may see this as happening through the negotiated and open debate over one element of statistical practice (namely monitoring and standardization), but where the wheel of counting, comparing and accounting practices keeps rolling with the drive of a parallel and not specifically relevant common authorization.

This paper does not address the question of why the strategic level wishes such audit and quantification, and this will be part of my further work. Without going into this question in depth, the answer may lie in a desire to combine several goals through one simple reporting system. These goals may be e.g. meeting public demands for transparency, enabling public roles as market customer, tracking providers' performance of contractual obligations, enforcing selected standards, performing management control functions from within a down-scaled management structure. It may be just because it is simpler and therefore a more powerful tool for administration regardless of the quality produced, or it may be because it is a matter of a method that has become a goal in itself; standardization and counting for standardization and counting's sake. These questions are left unanswered in this context. The

⁸ This paper is named by a triple reference. Latour introduced Pandora's black box (Latour 1987) to illustrate how scripts in technologies are hidden by black boxes. The cyber technicians use the concept of black box to figure out areas in the procedures where the operations are so complex that it is better to not involve more than what goes in and what comes out. Pandora's box, according to the myth, was not to be opened, or the all the sorrows of the world would escape and torment the world. Latour's conception is that this is how science and technologies work: The methods get established, and by reopening them we would find so many unsatisfying conclusions that were really just solidified by habit, and have to ask questions about the results that we would otherwise blindfold ourselves to. When I term it Rasputin's black box, it is to imply that this may also apply not only to inquiry of the scientific process, but also to processes of technologies of government.

Rasputin never did have great power, so the allegory is referring to the idea of Rasputin (the shadowy, discrete and powerful authority that is outside public reach), not of the historical person Rasputin (A troublemaker that was allowed to do as he pleased, as the Tsarina was absolutely devoted to him). The allegory is that of the invisible power, to which we can scarcely add faces or characters, because the meanings are "trapped" in black boxes of technology and established understandings.

conclusion of this paper is that we observe a mixing of standardization and quantification in statistics that serve to give the impression of a dialogue based consensus, when there is more to the way the statistics is used that is not included in this consensus. As this paper points to, matters of audit through quantification is definitely part of how the statistics are used, but probably not part of the consensus.

This conclusion does, as mentioned, presuppose a number of omissions, and they may be summed up in one question that should be further studied: What does this consensus (and non-consensus) that I address really contain? *Monitoring*, (not *audit*) is obviously one out of many ways to describe it, and other possibilities should be interrogated. Furthermore, there are quite certainly other processes as well that contribute to the growth of audit culture in organizations as these, and this paper does not intend to provide a full explanation of this, only to investigate one of the possible legitimizing drives. Finally, if one accepts monitoring and audit, there are possibilities for further investigation of the relationship between these, and the concepts of standardization and quantification; to which degree this is a fitting parallel, and what this connection would further contain.

REFERENCES

Cooke, R. 2005 *The history of mathematics: a brief course / Roger Cooke.*

Desrosières, A. 1997 'Du singulier au general. L'argument statistique entre la science et l'État', in I. B. C. L. Thévenot (ed) *Cognition et information en société*, Paris: Éditions de l'École des Hautes Études en Sciences Sociales, Translated by Linda Sangolt.

— 1998 *The politics of large numbers: a history of statistical reasoning / Alain Desrosières ; translated by Camille Naish.*

— 2001 'How Real Are Statistics? Four Possible Attitudes', *Social research* 68(2): 339-355.

Gigerenzer, G. (ed) 2001 *Bounded Rationality: The Adaptive Toolbox / Ed. G. Gigerenzer and R. Selten*, London: The MIT Press.

Hacking, I. 1990 *The Taming of Chance*, New York: Cambridge University Press.

Høvik, S. 2004 *Kommunal organisering 2004 : redegjørelse for Kommunal- og regionaldepartementets organisasjonsdatabase / Sissel Hovik og Inger Marie Stigen.*

Lampland, M. (ed) 2006? *Reckoning with Numbers. Forthcoming*

Latour, B. 1987 *Science in action: How to follow scientists and engineers through society / Bruno Latour.*

Porter, T. M. 1995 *Trust in numbers: The pursuit of objectivity in science and public life / Theodore M. Porter.*

Power, M. 1997 *The audit society: Rituals of verification.*

Strathern, M. 2000a *Audit cultures: Anthropological studies in accountability, ethics and the academy*, London: Routledge.

— 2000b 'New accountabilities' *Audit Cultures. Anthropological studies in accountability, ethics and the academy*, London: Routledge.