Draft presented at the Workshop on Local Governance Today: European and Turkish Experience in Local Politics, Democracy and Governance and Reciprocal Lessons. Ankara, Turkey, 20-24 March 2013
Alexander Kalgin
PhD candidate
Department of International Development, University of Birmingham, UK
Department of Public and Municipal Administration, Higher School of Economics, Russia
ask702@bham.ac.uk

Performance management in the Russian public sector: evidence of data manipulation at the regional level

Abstract

Performance management in the public sector appears to have become old-fashioned in the Western world. At the same time, non-Western countries appear enthusiastic with performance measures and indicators. Russia presents a good example. New initiatives appear on a regular basis. What drives this enthusiasm?

I find that performance management in Russia resembles the Western NPM technique in name but not in substance. The system has evolved to accommodate local needs, values and priorities of the "vertical line of power" in a centralized hierarchy.

The evidence comes from two sources: qualitative data was collected by interviewing 25 current and formal civil servants (including tree ministers of economic development) in 3 regional governments in Russia. Quantitative data comes from the publicly available performance dataset covering the period of 2007-2011. With more than 300 indicators and 83 regions this dataset provides more than 90 000 observations.

I find that compulsory data reporting facilitates data manipulation and misrepresentation. I demonstrate statistical evidence for two types of strategic behavior of bureaucrats subjected to formal performance reporting. Prudent bureaucrats minimize their long-term risks by reporting stagnant figures. Ambitious bureaucrats inflate figures to maximize their short-term career gains.

I demonstrate that performance indicators in the Russian public sector have little managerial value. They are seldom used to facilitate better decision-making or more results-oriented budgeting. Instead they have become an instrument of bureaucratic competition and upward accountability.

The results contribute to the wider literature on public sector reform. I stress the importance of local context, values and mentality in designing effective reforms. On the basis of this analysis recommendation are made to limit the incentives for data manipulation by introducing formal audit, ensuring independence of data collection bodies and focusing on wider reforms of public sector ethics.

Introduction

Although the problem of deliberate data manipulation has been long known to scholars of performance management both in private (Sims et al. 1987) and in the public sector (Smith 1995) it has scarcely been investigated quantitatively. So far evidence of deliberate data manipulation in performance management systems has been predominantly anecdotal and qualitative (for an excellent summary, see (Hood 2006)). Some attempts have been made to assess systematic data bias resulting from partisan interests (Gueorguieva et al. 2009), but so far I know of no attempts to capture systematic data bias that results from incentive structures inherently present in any bureaucratic hierarchy.

Guides for public sector practitioners rarely address the problem formally (Local Government Information Unit. 1994) and even well-known performance management gurus seem to work on the assumption that performance data may be accepted on its face value and be the basis for rational decision-making (Barber 2007).

At the same time, the danger of deliberate distortion of performance data by reporting agents is both theoretically predictable (Smith 1995) and practically observable (Hood 2006; Longenecker and Ludwig 1990).

I assert that particular patterns of deliberate data manipulation may be the direct result of civil servants' rational pursuit of their self-interest. The extent to which such patterns are pronounced will depend on the system of checks both formal and informal that prevents bureaucrats from engaging in data manipulation.

I use the dataset of Russian regional governments to demonstrate the patterns and argue that the particular conditions of the Russian civil service are favourable for the observation of this phenomenon.

The following text is organized as follows: first, potential motives that may lead to data manipulation are outlined; second, Antony Downs' (1967) theory of bureaucracy is used to construct models of bureaucratic behavior towards manipulating data; third, the methodology of the original research is presented; fourth, the background of Russian public sector reform is briefly portrayed; fifth, the dataset is described and the analytical framework is laid out; sixth, the results of the analysis are presented and, finally, some wider implication of the results are discussed.

Motives behind deliberate performance data manipulation

Substantial understanding of motives behind deliberate performance data manipulation has been gained by researchers of performance management systems in the private sector. Longenecker and Ludwig (1990) provide a useful overview of lessons learned. In the 1980s there was a significant interest among businesses in implementing formal performance appraisals. Modern performance management techniques in the public sector have much in common with these earlier tools. Both systems generate similar ethical problems. Both systems are susceptible to deliberate data manipulation:

> *In theory, … employee appraisal is an objective, rational, and systematic attempt on the part of the manager to accurately describe subordinate performance. In reality, however, managers have a variety of concerns that are clearly more pressing than simply generating brutally accurate and honest ratings. (Longenecker and Ludwig 1990: 962)*

The chief reason for this is that managers are driven by their self-interest, they 'operate in organizational environments that place high priority on getting results, on minimizing conflict, and ultimately, on survival' (Longenecker and Ludwig 1990: 962). Producing a 'brutally accurate' performance rating may not be the best course of action in this case.

A number of reasons may lead managers to distort performance information. They may both inflate and deflate performance figures. Longenecker and Ludwig (1990: 962) provide a typology of motives that may lead to such behavior. These motives may be either positive (the manager may believe that he is acting in the name of some higher purpose) or deviant (driven by pure self-interest).

Table 1. A typology of rater motives and manipulative rating behavior (Longenecker and Ludwig, 1990:966)

Rater's motives	INFLATED RATINGS	DEFLATED RATINGS
POSITIVE	• Keep the employee motivated • Maximize the merit pay increase • Avoid creating a permanent record that might damage the employee's career • Reward good recent performance • Assist an employee with a personal problem • Reward effort • Liking the subordinate personally	• Scare better performance out of an employee to prevent eventual termination • Build a stronger case against an employee who is destined to be terminated
DEVIANT	• Avoid hanging dirty laundry in public • Make themselves look good • Avoid conflict/confrontation — with a subordinate • Promote a problem employee up and out	• Punish an employee • Encourage an employee to quit • Minimize merit pay increase • Comply with an organizational edict to keep ratings low

With minor alterations this typology may be applied to explain motives behind deliberate data manipulation in the public sector. The three priorities attributed to private sector organisational environment are similarly applicable to the public sector. Getting results, minimizing conflict and, ultimately, survival are key priorities for any manager.

In order to adapt the typology to the public sector and estimate the relative importance of different motives for civil servants a theoretical understanding of bureaucrat's utility function is needed. Such understanding is provided by Antony Downs' seminal work on bureaucracy (1967).

Bureucrat's utility function

The "utility functions" of bureaucrats according to Downs (1967 : 85) are made up of self-interest and altruistic (broader) motives:

Figure 1. Motives of bureaucrats according to Downs (1967) (Dunleavy, 1999, p.148-9)

Self-interest motives	*Broader motivations*
• *Power* - inside the bureau or outside it. • *Money income*. • *Prestige*. • *Convenience* - minimizing personal effort. • *Security* - defined as a 'low probability of future losses of power, money income, prestige or convenience'.	• *Personal loyalty* - to the immediate work-group, bureau as a whole, the wider government, or the nation. • Identification with a specific program of action or "*mission-commitment*" • *Pride in proficient performance of work*. • *Desire to serve 'the public interest'* - that is, what the official believes the bureau should be doing to carry out its social function

Downs identifies 5 bureaucratic personality types corresponding to these motives: two "purely self-interested" and three "mixed-motive" types (Downs 1967: 88):

Figure 2. Personality types of bureaucrats adapted from (Downs, 1967, p.88)

Personality type	**Dominating motives**	**Comment**
	Purely self-interested officials	
Climbers	power, income, and prestige	
Conservers	convenience and security	In contrast to climbers, conservers seek merely to retain the amount of power, income, and prestige they already have, rather than to maximize them.
	Mixed-motive officials	
Zealots	mission-commitment	They seek power both for its own sake and to effect the policies to which they are loyal.
Advocates	personal loyalty	They are loyal to a broader set of functions or to a broader organization than zealots
Statesmen	desire to serve 'the public interest'	They are loyal to society as a whole, and they desire to obtain the power necessary to have a significant influence upon national policies and actions.

Bureaucrats belonging to different personality types have different value systems and rank their motives differently. This difference may result in resorting to different courses of action when an opportunity comes to engage in data manipulation. Individual decisions of civil servants, if made systematically, generate certain patterns in resulting statistics.

Reflection of bureaucratic personality types in performance data

In order to illustrate the process that leads to systematic bias in overall performance data, it is useful to consider a hypothetical situation where a bureaucrat is required to report performance figures to his superiors with no provisions for audit of the data.

Observations made during the qualitative stage of the research allowed me to create models of behaviour for two self-interested types of Downs' bureaucrats: the conservers and the climbers.

Models of unaudited data reporting

In a hypothetical situation where there is no routine audit of the reported data and the "real-life performance" is stagnant the two self-interested types of bureaucrats are likely to adopt different strategies in respect to data manipulation.

Conservers aim at minimising risks and maintaining *status quo*. They prefer to play safe and, if they find themselves in a situation where no one routinely checks the validity of their bureau's reported performance data, they resort to the "prudent bureaucrat" strategy. They report minor variations in the data as this is least likely to cause suspicion and attract attention of their superiors. No one can be blamed for a drop of 1% in annual performance, and a 1% annual growth is also inconspicuous. As one civil servant observed:

> *If you inflate your figures too much, people may start asking questions. So, you shouldn't be too thin and shouldn't be too thick. You have to take the prudent approach.*

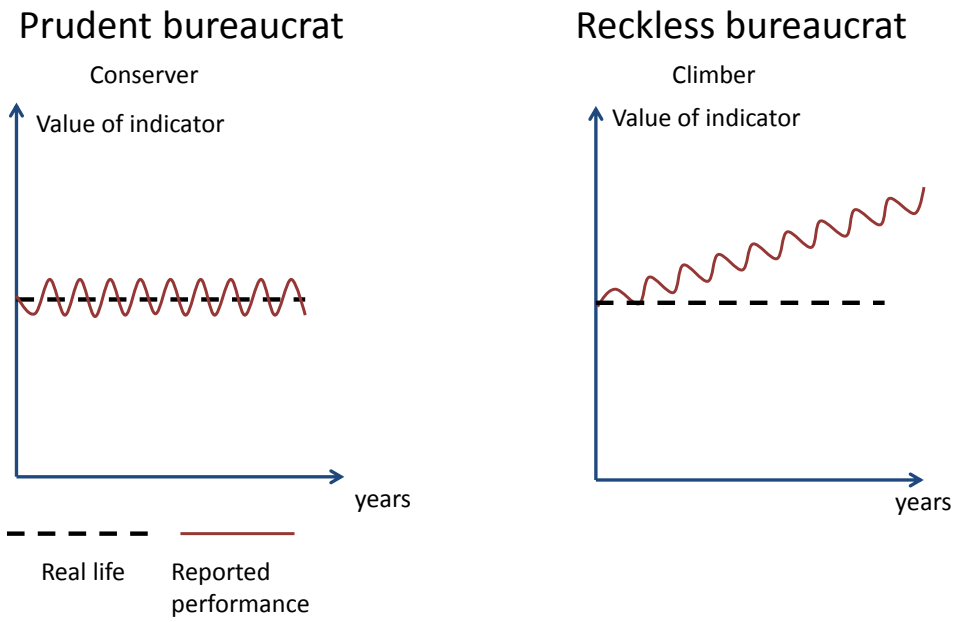
Climbers aim at maximizing their career prospects, thus, they attempt to show their agencies' performance in favourable light no matter what. If faced with the same situation, they are likely to engage in deliberate inflation of their performance figures. One lower level civil servant gave the following specific example:

> *When I worked in the Ministry of culture, we had to show that we have a growing number of social clubs in rural areas. But there was no money and the people were leaving, so, actually, the number of social clubs was going down. But we had to demonstrate growth and we did. Year after year*[1].

These types of behavior are illustrated in Figure 3.

[1] In this example, in fact, the decision to inflate performance figures was taken out of the fear of being sanctioned for poor performance and, thus, may be considered a conserver's coping strategy. The same course of action, however, would be taken by one who was looking to please his subordinate out of the desire for a promotion. Whatever the motive, however, the resulting behavior results in unrealistic inflation of the data and, thus, may lead to severe sanctions if the fraud is found out.

Figure 3. Strategies of data reporting adopted by climbers and conservers in the absence of routine audit.



It is prudent from a *conserver's* point of view to avoid inflation of performance figures. A conserver is motivated by the desire of maximum long-term security and minimum inconvenience. Thus, it is preferable for him to be inconspicuous. Any above-the-average as well as below-the-average performance is risky as it may attract superior's attention.

In contrast, a *climber* that engages in reckless inflation of his performance figures prefers to take risks. If he is caught he is likely to suffer severe penalties, but if he successfully conceals his fraud he may earn a promotion or obtain other perks.

Using evidence from the national performance dataset I demonstrate that among Russian civil servants at the regional level "prudent bureaucrat" strategy dominates but traces of "reckless" behaviour are also observable once additional perks in terms of grants are introduced.

Background: performance measurement in the public sector in Russia

> *The 1990s were not kind to public service reform in Russia. (Barabashev and Straussman 2007: 379)*

The history of public sector reforms in modern Russia may be subdivided into 5 distinct stages. Barabashev and Straussman (2007) identify 4 stages in the period between 1992 and 2006 and (Jakobson 2001) adds another transition stage in 1990-1992.

These stages include: the first stage of turbulence immediately after the collapse of the Soviet Union, when the raging political struggle meant that the executive functioned without well-defined and certain leadership (1990-1991) (Jakobson 2001); the three following preliminary stages that established the structure and defined responsibilities of federal executive bodies thus setting the scene for a systematic

reform effort (1992-2003); the contemporary stage during which a consistent reform process began (2003-present) (Barabashev and Straussman 2007).

Currently public sector reforms are organised in three separate strands: the civil service reform, the budget reform and the administrative reform (Verheijen and Dobrolyubova 2007: 210). The civil service reform aims at developing a professional civil service by creating incentives for performance at the individual level. The budget reform aims at improving the budget process and inter-budgetary relations. The administrative reform aims at optimising the structure of government and streamlining its operations (the major achievement of this reform has been the introduction of a new three-tier government structure at the federal level (Ministry-Service-Agency).

Performance management initiatives have been mainly implemented within the framework of the administrative reform. Commentators observe that so far the principles of performance measurement have not been adequately reflected in the two remaining strands of reforms (Verheijen and Dobrolyubova 2007). Thus, performance measurement mechanisms in inter-governmental relations have not been synchronized with corresponding initiatives in performance budgeting or individual performance-related pay.

Nevertheless, a nation-wide system of performance measurement was established in 2007 and has since been evolving[2].

The most significant and consistent effort in performance data collection was initiated by the Decree 825. According to this decree all 83 regional governments of the Russian Federation were required to collect and publish annual performance figures for a set of nearly 300 indicators. The system operated with minor alterations between 2007 and 2011[3]. In 2012 the dataset covering the 5-year period was compiled and published by the Ministry of regional development. This dataset has provided opportunity for systematic research of trends in public sector performance over time. It is particularly valuable as it allows for quantitative testing of hypotheses about deliberate data manipulation. I argue below that existing conditions have created a situation where deliberate data manipulation is not only unrestrained but is, in fact, encouraged.

[2] The foundations of the current system were laid by three main pieces of legislation:
Presidential Decree of 28.06.2007 № 825 "On the evaluation of performance of the executive authorities of the subjects of the Russian Federation" (Decree 825);
Presidential Decree of 28.04.2008 № 607 "On the evaluation of performance of local governments in urban districts and municipal areas" (Decree 607);
Presidential Decree of 21.08.2012 № 1199 "On the evaluation of performance of the executive authorities of the subjects of the Russian Federation" (Decree 607).

[3] In 2012 the system was radically modified: the number of indicators was reduced from over 300 to under 60 indicators.

Methodology

Sample

The research was conducted in two stages: first, 25 semi-structured interviews were carried out with officials from regional governments, consultants and academics[4]. Motives and opportunities for deliberate data manipulation were identified and hypotheses regarding their manifestation in resulting performance data were formulated. Second, quantitative evidence was obtained by analysing the nation-wide performance dataset[5] that provides figures for over 300 performance indicators covering 83 regional governments for the period of 5 years between 2007 and 2011 (this gives over 100 000 observations).

The dataset

The data of the national dataset are used by the Ministry of regional development to allocate grants to regions. At the same time it is also used by the presidential administration to evaluate governors[6] and by governors to evaluate their subordinates. Regional administrations use the same data to evaluate municipal civil servants. Thus, a pyramid of principal-agent relations in created within the "vertical line of power". The same data are used multiple times by multiple principals to evaluate their corresponding agents.

Performance data in the national dataset are provided by various government agencies. The dataset is centrally compiled by the Ministry of regional development, but values of individual indicators originate from different sources. The complete structure of the dataset is given in Appendix 1.

Reporting agencies were grouped into four groups according to the model of principal-agent relationship that most accurately describes their interaction in relation to data collection, reporting and evaluation. 3 such models were identified: A) impartial reporting agency; B) interested reporting agency; C) self-reporting by the agent.

Models of principal-agent relationships

Model A – Impartial reporting agency

In this model the function of data collection is performed by a third party that is not affiliated directly with either the principal or the agent. In the Russian case this model operates where the data is collected by either the Federal Bureau of Statistics (Rosstat) or the Federal Security Service (FSO).

[4] The sample included: 16 current regional civil servants, including three ministers of economic development, three deputy ministers from departments of health and economic development, 6 heads of departments and 2 specialists (lower level civil servants); 9 former civil servants, including: 3 consultants, 3 academics. Other respondents included: a top central government official responcible for the design of the nation-wide system of performance management and two civic activists involved in public scrutiny of government performance reports. The interview guide from (Pollitt 2006) was used with minor alterations. Purpose sampling and snowballing were used to generate the list of respondents. People responsible for implementing and operating performance measurement systems were identified in regional governments and asked to provide further contacts for interviews.

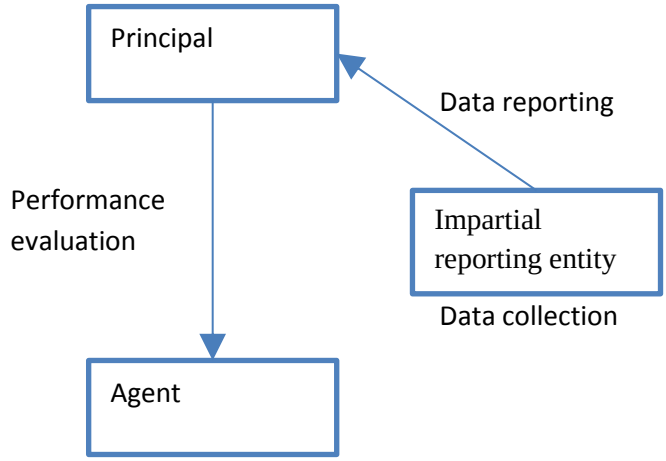
[5] The data was collected in accordance with the Presidential Decree № 825 of 28.06.2007 "On assessing the effectiveness of executive authorities of the subjects of the Russian Federation.", The data was published by the Ministry of regional development in 2012.
URL: http://www.minregion.ru/upload/documents/2012/10/101012/101012_itogi_2011.xls [accessed 08.03.2013]

[6] Governors of Russian regional governments have been appointed by the president since 2004. There are no formal rules on what is considered good performance and what career decisions would follow if performance indicators demonstrate good or bad dynamics. The system relies on the set of informal incentives to generate competition among governors (Libman et al. 2012).

Neither of these two bodies is evaluated on the grounds of performance data that they provide. It is not in the interest of federal statisticians to inflate or deflate figures for any particular indicator or a particular region. They may be considered reasonably impartial[7]. Similarly, federal security agents are deemed to be above the game[8].

In this model there is assumed to be little or no incentive for deliberate data manipulation.

Figure 4. Model A - Impartial reporting agency



Model B – Interested reporting agency

In this model the reporting agency may have an indirect interest in demonstrating desirable trends in the data. In our case this model operates where a federal line ministry (such as the Ministry of Education) is responsible for collection and reporting the data on education performance measures.

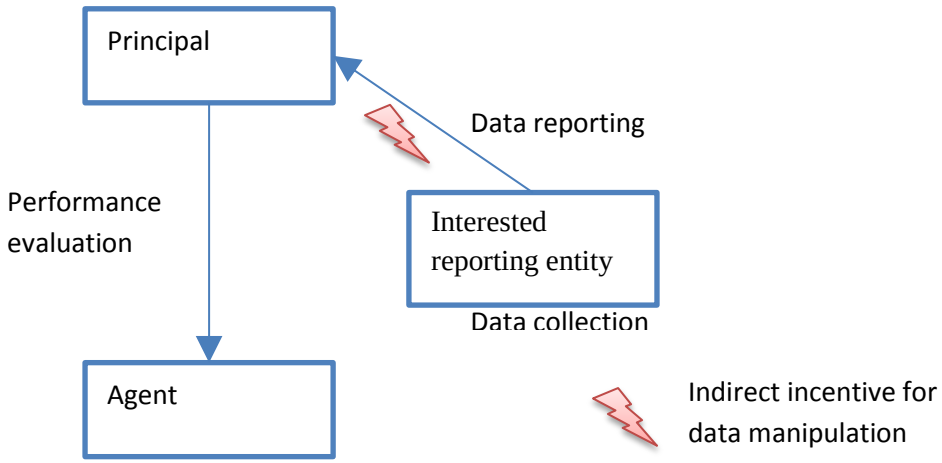
In this model there might be an indirect incentive for data manipulation. The reporting entity is at the same time responsible for government policy in the area under consideration and is, therefore, indirectly assessed on the basis of data that it collects. Thus, a conflict of interest may occur. The agent, however, may be motivated to check the data to ensure its validity and fairness[9].

[7] A recent confirmation of this may be found in the fact that the Federal Statistics Bureau (Rosstat) recently published population estimates that contradicted political statements of Prime Minister Medvedev. The politician announced that in 2012 the population of Russia demonstrated positive growth for the first time in decades and credited the government for this achievement. A few days later Rosstat published the results of annual survey that showed a minor drop in total population. This contradiction was commented on as a sign of political impartiality of federal statisticians. (Gontmacher, 2013, http://echo.msk.ru/blog/gontmaher/1003510-echo/).

[8] The Federal Security Service (FSO) is charged with collecting citizen satisfaction data. One civil servant commented on this fact: "*When we used to collect citizen satisfaction figures in the Department of Health, we used to get 90% satisfaction all the time. Once FSO took over, they started showing 35%".*

[9] If, for instance, the Ministry of Education were to engage in deliberate distortion of performance figures it would have to attribute higher ratings to some regions and lower ratings to other regions and such discretionary changes could cause dissatisfaction with regional administrations. It seems unlikely that such manipulation could be widespread, but it was not deliberately investigated in this research project.

Figure 5. Model B – Interested reporting agency

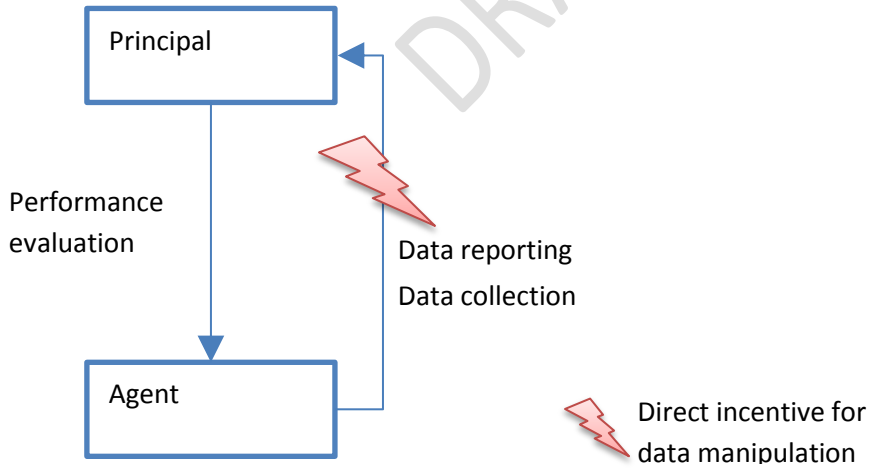


Model C – Self-reporting by the agent

In this model the data is collected and reported by the agent. No independent body is involved. In our case this model operates in the situation where regional governments collect, analyze and report data on their own performance.

In the system that operated between 2007 and 2011 there were no provisions for formal routine audit of data collected in this way. When asked whether they consider such data reliable many civil servants were puzzled and were not able to give an answer. Most respondents thought that "there must have been somebody who checked the data", but were unable to give details[10].

Figure 6. Model C – Self-reporting by the agent



[10] One regional minister of economic development was clear that he did not trust locally generated data: "*However well qualified local civil servants are, one should never rely on their data. The growth of any indicator depends on the person who seats in the courtyard of his collective farm and counts. One should only rely on the data of the Federal Bureau of Statistics".*

Grouping of reporting agencies

According to the most appropriate model of principal-agent relationship the agencies that contribute to the national performance dataset were categorised into four[11] groups: 1) Federal Bureau of Statistics; 2) Federal Security Service; 3) Federal Line Ministries; 4) Regional Administrations. The total number of indicators that are reported by each group is given in Table 2.

Table 2. Number of indicators in groups of agencies

Group	Corresponding model of principal-agent relationship	Number of indicators
Regional governments	Model C – self-reported	72
Federal Bureau of Statistics (Rosstat)	Model A – independent agency	75
Federal line ministries	Model B – interested agency	174
Federal Security Service (FSO)	Model A – independent agency	8

Hypotheses

Assuming that incentives to manipulate the data result in distortions in the resulting figures, the following hypotheses were formulated:

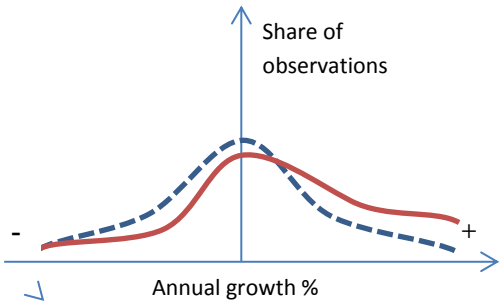
H.0 – Indicators collected by all four groups of agencies demonstrate similar behavior over time.

H.1 – Indicators collected by four groups will demonstrate different dynamics over time.

I assume that Model C (self-reported) allows bureaucrats to engage in deliberate data manipulation with least constraint. Thus, in this case both *climbers'* and *conservers'* behavior will be more pronounced and observable in the data.

H.2 – Self-reported indicators will demonstrate more positive dynamics over time. (Climbers inflate the data to obtain perks) (Figure 7[12]).

Figure 7. Predicted distribution of annual growth indices for "reckless" data manipulation

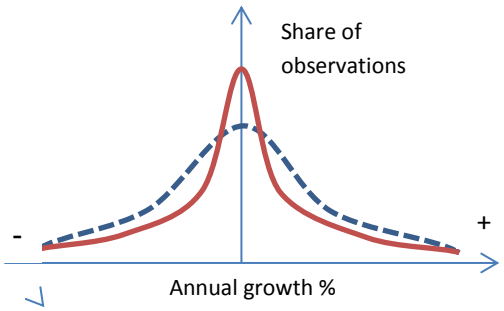


[11] I categorize the Federal Security Service (FSO) into a group of its own due to the peculiar nature of indicators that are assigned to this body. FSO collects citizen satisfaction data for all areas of governance (total of 8 indicators). The fact that satisfaction indicators were assigned to this organization may be interpreted as evidence of the low level of confidence in data provided by line ministries and regional administrations.

[12] On Figure 7 and Figure 8 the blue dashed line represents the "normal distribution" of annual growth indices in a hypothetical situation where "the real life performance" is constant. In such a situation annual growth rates will be distributed in a way bearing resemblance with the normal distribution. Most regions would demonstrate insignificant growth/reduction with a minor share of outliers on both sides. If, however, there is a deliberate effort to avoid reporting negative figures (for climbers), or to demonstrate average performance (for conservers), the distribution will be skewed. The analysis has much in common with techniques used to capture effects of electoral fraud (see, for example, (Enikolopov et al. 2012))

H.3 – Self-reported indicators will be more stable over time. The distribution of growth indices will be crowded near zero values. (Conservers prefer "prudent bureaucrat" behavior and aim to be inconspicuous by demonstrating stagnating performance figures) (Figure 8).

Figure 8. Predicted distribution of annual growth indices for "prudent" data manipulation

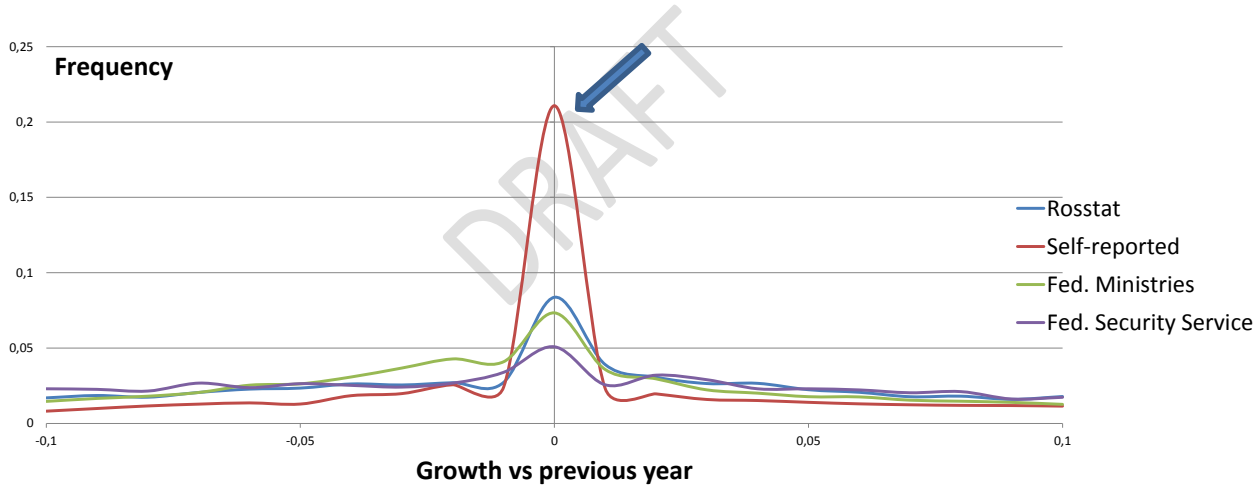


Results

Prudent bureaucrat's behaviour

The analysis of annual growth indices for indicators from the four groups demonstrates that self-reported indicators exhibit significantly different behavior from the three other groups. As predicted by H.3, self-reported indicators are clustered around zero to a much higher extent. This suggests that regionals officials could, in fact, adopt the "prudent bureaucrat" approach. Significantly larger proportion of observations in this group demonstrates near-zero growth (Figure 9).

Figure 9. Distribution of annual growth-indices[13] of indicators in four groups of agencies in the period of 5 years



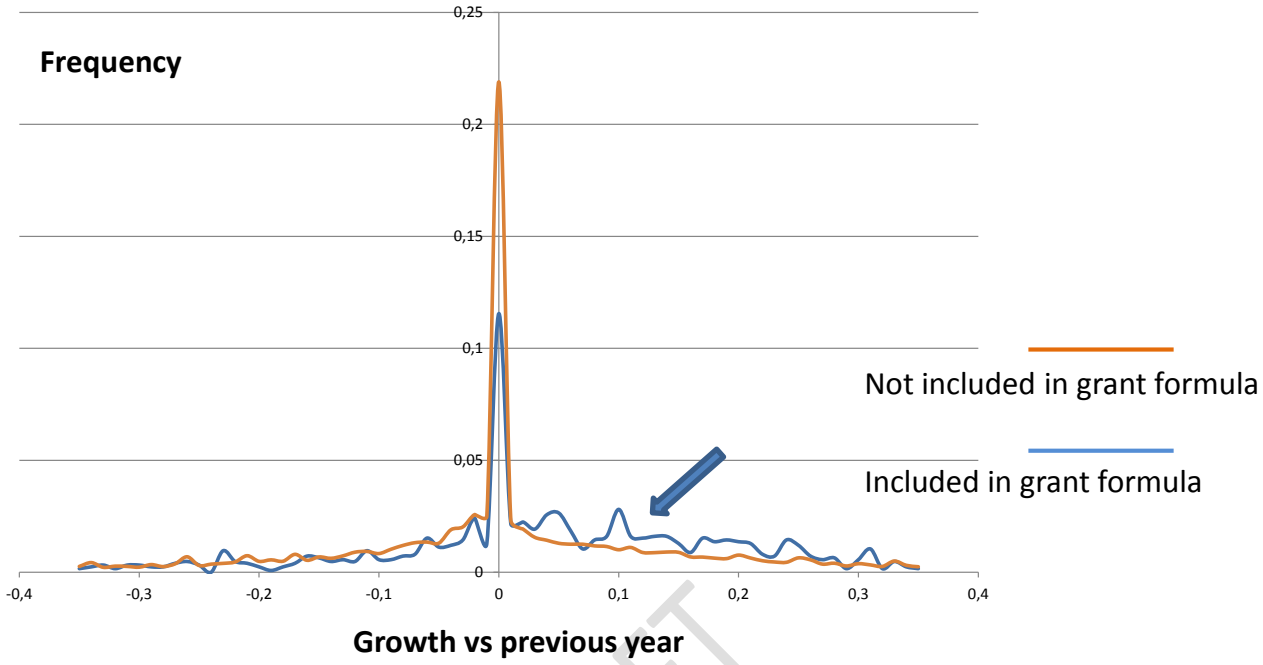
Reckless bureaucrat's behavior (stimulated by grants)

It is not evident from the analysis of the entire set of indicators whether the "reckless bureaucrat" behavior is unambiguously present. The right-hand tail of the graph is not perceptibly greater than the left-hand tail, thus significantly better overall performance is not detectable. This may suggest that, in general, prudent behavior prevails.

[13] Growth indices are calculated using the following formula: $\frac{PI_t - PI_{t-1}}{PI_{t-1}}$, where PI_t and PI_{t-1} – values of a performance indicator in years *t* and *t-1*, respectively. The lines demonstrate frequencies of occurrence of a given value as share of the total number of observations in this group. Number of observations (N) for the four groups: Rosstat – 22186; Self-reported – 15866; Federal Ministries – 48563; Federal Security Service – 2566 (4 time periods, 83 regions, number of indicators in a group - Table 2).

However, if one considers the subset of self-reported performance indicators and differentiates between those that were and were not included in the formula for intergovernmental grants a different picture may be observed Figure 10:

Figure 10. Distribution of growth indices of self-reported performance indicators (the effect of grants)[14] over the period of 5 years



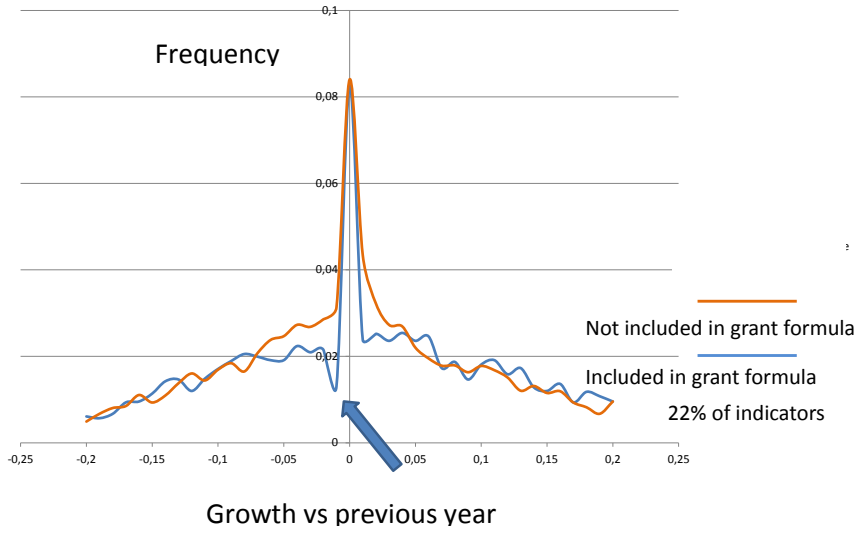
Self-reported indicators included in the grant formula for federal transfers demonstrated significantly better dynamics. A possible explanation is that by providing additional perks for positive dynamics, grants stimulate governors to pay attention to values of these indicators. This can lead to both increased real activity of subordinates to improve performance, or create greater incentives to inflate the figures. However, due to the low number of observations these results may not be conclusive.

Somewhat surprisingly, it turned out that the effect of grants is also pronounced in the group of indicators reported by the Federal Bureau of Statistics. A major drop in the left hand-side may be interpreted as evidence of the reluctance of reporting agencies to demonstrate minor negative figures in indicators that are included in the grant formula[15] (Figure 11).

[14] Out of 72 self-reported indicators only 7 were included in the grant formula giving $N_{included} = 1247$; and $N_{not\ included} = 14619$ (the total number of indicators included in the grant formula from all four groups was 69).

[15] The collection of data may, in fact, be performed by the same people for different reporting agencies. The difference, however, is in the number of layers of "interested bureaucrats" that the data has to go through before it appears in the final report. Front-line data collectors may not be in the position to collude and cooperate to systematically distort the data in one pre-meditated way. But once the raw data is collected it has to be processed and prepared for publication. There may be numerous revisions by different people and the resulting distortion depends on the propensity of each person to "massage" the figures. It may be that even data reported by Federal Bureau of Statistics is subject to such distortion at some point in the process.

Figure 11. Distribution of growth indices in the group of indicators reported by Rosstat (the effect of grants) over the period of 5 years.[16]



Discussion

The discussion of unintended consequences of performance measurement in the public sector has been going for long. Smith (1995) identified 8 types of unintended consequences that result from introducing formal performance measurement. Some of these are related to "real" physical effects on behavior such as concentrating on measured aspects of a service at the expense of unmeasured ones. Other unintended consequences are related solely to data generation process: performance may be misrepresented in various ways. However, these assertions tend to be supported exclusively by anecdotal evidence. Quantitative methods are rarely employed to demonstrate the scope and pervasiveness of data manipulation. Partly this is due to the difficulty in separating effects of deliberate data manipulation from effects of other factors. I suggest that the system of performance measurement that operated in Russian regional governments between 2007 and 2011 allows for quantitative analysis because of some of its unique features: 1) public availability of the data; 2) existence of clearly definable groups of reporting agencies with varying incentives to manipulate the data; 3) no provisions for routine formal audit; 4) no provisions for formal sanctions for data manipulation; 5) the "hermetic" nature of modern Russian bureaucracy that insulates it from public scrutiny and largely from political interference at local and regional levels.

Hermetic bureaucracy as a natural laboratory

The data provided by Russian national performance dataset has immense advantages from academic point of view. The data are collected by different agencies creating nearly "natural experiment" conditions and allowing to isolate the effect of the varying types of principal-agent relations on reported data.

The parameters of the system create favourable conditions for the observation of behavior of unaudited bureaucrats. It may fairly be said that the way in which the system was implemented and has operated may reasonably be seen as allowing and, perhaps, even encouraging data manipulation. Longenecker and Ludwig (1990: 968) argue that in order to be in the position to expect accurate performance reporting from managers the supervising organization has to fulfill a number of obligations (these

[16] Number of indicators reported by Rosstat – 75, out of which 14 were included in the grant formula. $N_{included}$ = 4918; $N_{not\ included}$ = 17268

suggestions originally relate to performance appraisals of individuals but are equally applicable to performance evaluation of organizations):

- *[The] organization must provide **a sound procedure** for managers to use in the execution of performance appraisal.*
- ***Training** in performance appraisal must be provided for managers. This training should formally address the issue of intentional inaccuracy, and should deal not only with manager's ability, but willingness and motivation to execute accurate ratings)….*
- *Organizations must provide **leadership from above**. Middle and lower level managers cannot be expected to provide accurate ratings if intentional inaccuracy is practiced higher in the organization)….*
- *Like any procedure, the performance appraisal **process must be audited** to ensure the accuracy of the data generated and the overall integrity of the process...*

It became evident during the interviews that none of these conditions was observed in the implementation of this particular system.

Another reason why the Russian case may be considered exemplary is the exceptionally high degree to which the Russian system of civil service is insulated from external scrutiny. The term "hermetic bureaucracy" has been used to describe it. Neither the civil society, not politicians have been able to establish effective control over public sector bureaucrats. This is generally considered a great flaw of the Russian system of civil service, but it is a great advantage for the academic research of bureaucracy as it allows us to observe pure bureaucratic behavior, untarnished by interventions of other systems.

One regional minister of economic development summarized his experience of dealing with the system of performance measurement as follows: "*I have a feeling that these are just documents for documents' sake… they are not linked to budgeting… their managerial value is absolute zero".*

For academic purposes, however, they have very high value as they allow one to observe strategies of bureaucratic struggle for survival and analyze them separately from influences of managerial decision-making, political process or demands of the public.

Performance data as means and as an end in itself

Longenecker and Ludwig (1990) make an important distinction between the use of performance data as an end in itself (this is the way the data are used by Human Resource departments in private organisations or by performance management units in governments) and as means to obtain other ends (this is the way performance data are used by managers or public servants who generate it). In the first case, the data is valuable because it provides the material for decision-making by the principal; in the latter case performance data is used to achieve some goals other than higher performance in its narrow sense. Managers may use performance data to formally justify rewards to selected subordinates or to provide official backing for politically motivated decisions. Performance figures are then just by-products of a wider activity. Whatever the reason, the result is that accuracy of performance data is undermined.

Moreover, from a public sector manager's point of view the very fact that performance data is ambiguous may be an advantage. Ambiguous figures may be used to rationalise discretionary decisions. As one consultant observed:

Among three hundred indicators the governor can always find those that can be used to either reward a minister or to dismiss him.

The natural selection in the "vertical line of power"

In his influential work on the theory of bureaucracy published in 1965, Gordon Tullock (2005) used the analogy of gaseous diffusion plant to illustrate the process of career selection in any hierarchy based on merit. Individuals entering the system are continuously tested and either rise, fall or remain in place as a result of the test. Only those who systematically make decisions in the interest of their career are likely to rise to the top:

> *Any political hierarchy in which personnel are selected for promotion by the system we have designated "merit" will function in much the same way. People entering the system are either a random selection or the result of a preliminary selection process. Once they are in the system, they are confronted with a number of situations in which they may either rise, remain in the same position, or fall. These "test" situations do not necessarily refer to formal promotions in the bureaucratic hierarchy. There are usually numerous smaller steps which prepare the way for formal promotion or demotion. The obtaining of a good assignment, earning the confidence of your superiors, getting a "good name around the office," all may be equated to the porous barriers of the gaseous diffusion plant. (Tullock and Rowley 2005: 21)*
>
> *…*
>
> *Any individual in any system will continually be confronted with choices between courses of action which will have at least some favorable effect on his chances and others which are less desirable from that point of view, but which have other advantages. Only the person who usually chooses in terms of his "career" will be likely to rise to the top.(Tullock and Rowley 2005: 22)*

Since people can adjust their behavior to the rules of the game and since individual objectives often differ from organizational objectives, such a system inevitably selects against honest individuals and promotes those who always pursue the course of action that is favourable for their personal advancement even at the detriment of organizational objectives:

> *It is impossible to design a system that will select against the man of relatively low morals. This is because the intelligent but unscrupulous man will always assume the morally proper course of action if, in fact, this should be the one that is the most likely to be successful.(Tullock and Rowley 2005: 26)*

An "idealistic" individual who attempts to pursue "the general good" will find himself in a disadvantage when confronted with a choice between two actions one of which is likely to bring him a career promotion but is bad for the organization as a whole, and the other one which is in the true interests of the organization but is likely to hurt the individual. Thus, "any organizational structure in which selection on a merit basis is employed is likely, at least to some extent, to select against morality". (Tullock and Rowley 2005: 26)

A "realistic" individual, who habitually chooses in his individual interests, may not necessarily be aware of the damage he is inflicting:

> *It is always difficult to distinguish between "what is good for me" and "what is good." The general good is never readily discernible. The "politician," the bureaucrat, who makes no especial effort to keep these two categories*

distinct can quite genuinely believe that a course of action which may appear cold-blooded and dishonest to the outsider falls legitimately within his range of duty. (Tullock and Rowley 2005: 27)

In an ideal organization one never has to choose between "what is good for me" and "what is good". It is unlikely, however, that his situation is common:

In the ideally efficient organization, then, the man dominated by ambition would find himself taking the same courses of action as an idealist simply because such procedure would be the most effective for him in achieving the personal goals that he seeks. At the other extreme, an organization may be so badly designed that an idealist may find it necessary to take an almost completely opportunistic position because only in this manner can his ideals be served. The idealist, in such cases, may find that only by taking the course of action that will advance his own career can he remain in the organization and advance to a position where he can hope to influence events. This is administrative organization at its worst (Tullock and Rowley 2005: 24)

…

The general "moral level" of those bureaucrats who have reached the top layers in such a structure will tend to be relatively low". (Tullock and Rowley, 2005:25)

The public sector is known to be the place where one inevitably faces such dilemmas. Thus, naturally moral people may prefer to avoid public sector employment deliberately. A mechanism of negative self-selection is then set in motion:

Recognizing the dilemma with which they are likely to be faced in this respect, many highly perceptive and moral persons deliberately avoid employment in such hierarchical systems. Such persons recognize that they cannot be, by their nature, sufficiency dull as to remain subjectively honest in genuine conflict situations while they are unwilling, on moral grounds, to adopt consciously dishonest positions. In any event, few people expect career civil servants to act contrary to their own interests.(Tullock and Rowley 2005: 27)

This concluding remark brings us back to the topic of data manipulation. As Tullock observes, "few people expect career civil servants to act contrary to their own interests". This is particularly the case for Russian civil servants. It has been well documented that the problem of negative selection at the entry level is pervasive in Russia. Rent-seeking remains a significant motivation for those choosing to work as civil servants.

The situation is exacerbated by the dominance of informal rules and ambiguity in criteria for career promotion (Яковлев 2010: 24).

A recent survey of civil servants[17] found that over 73% think that the most relevant criteria of career promotion is the disposition of their immediate supervisor[18] and a large proportion of civil servants find criteria of career growth unclear and unpredictable.

It is, therefore, reasonable to expect that any ambitions civil servant when confronted with the choice of whether to manipulate the data would choose to please his immediate supervisor by showing figures

[17] http://cinst.hse.ru/contract (publications forthcomming)

[18] One of the questions read: "Who, do you think, is in the position to most accurately judge the results and quality of your work*?".* Responses: 73% immediate supervisor; 25% the head of department.
Another question: "How clear and predictable are the criteria of career growth in the civil service?" Responses: clear and predictable – 19%, all depends on the immediate supervisor – 24%, all depends on the head of department – 21%, criteria are unclear and unpredictable – 19%, do not know – 17%.

that show him and his department in a more favourable light. Since the criteria of career promotion are unclear and all depends on the discretion of supervisors, it is rational to try to portray a good image "just in case":

> *The paradox is that if one tries to introduce formal indicators without clearly defining their role and status, a system of informal incentives spontaneously emerges (including the desire to get noticed, etc). The resulting effect of such indicators on the system is hard to predict, account for and correct. (Якобсон 2006: 19)*

At the same time, when facing with a choice of whether to report a discovery of unrealistically inflated figures (perhaps, done by his predecessor or by another agency), it would take exceptionally courage and moral fiber to blow the whistle and call for an investigation. One who decides to uncover such manipulations may expect his career prospects to be worsened.

If this situation is present, the system may become self-destructing in the long run:

> *…the barriers act so as to select by criteria that are not only irrelevant from the standpoint of the designers, but which will, in the future, result in even poorer performance and selection.(Tullock and Rowley 2005: 24)*

If civil servants are allowed to manipulate the data unchecked, the whole edifice of performance management is likely to collapse as the data becomes increasingly detached from reality.

It is not certain whether the current system is prone to such extreme deficiencies, but it is informative to consider the fact that in 2012 the system of performance measurement in Russia was radically changed. Out of more than 300 indicators only 60 remained and out of these 60 not a single indicator was to be self-reported by the regional administrations.

It appears that the presidential administration has learned the lesson and discontinued the bulk of indicators that became largely meaningless. One factor of this could be the two data manipulation strategies outlined above.

Conclusion

Performance management systems are generally based on the assumption that performance data is collected objectively and accurately. Thus, it is assumed that rational decisions may be made on their basis.

It has, however, long been acknowledged that considerations other than "brutal accuracy" may enter in the decision-making process of managers who collect and report the data.

This article suggests a simple quantitative way of demonstrating the effects of deliberate data manipulation. The dataset generated by the Russian system of performance measurement at the regional level is used.

I find that compulsory data reporting may encourage data manipulation and misrepresentation. I demonstrate statistical evidence for two types of strategic behavior of bureaucrats subjected to formal performance reporting and suggest two possible explanations for such results. Prudent bureaucrats may choose to minimize their long-term risks by reporting stagnant figures. Ambitious bureaucrats may inflate figures to maximize their short-term career gains.

If no provisions are made to check such behavior, the usefulness of the system of performance measurement as a nation-wide tool of benchmarking and administrative decision-making may be undermined.

Appendix 1

Table 3. The sources of pefromance data in the national dataset.

Agency providing the data	Number of indicators
Federal Bureau of Statistics (Rosstat)	75
Regional Administrations	72
Federal Ministry of Health	47
Federal Ministry of Education	37
Federal Treasury	31
Federal Ministry of Regional Development	14
Federal Security Service (FSO)	8
Federal Ministry of Sport	8
Indicators that were re-assigned during the period	8
Federal Service for Control in Education (Rosobrnadzor)	7
Federal Ministry of Finance	6
Federal Ministry of Internal Affairs	5
Federal Ministry of Natural Resources	4
Federal Ministry of Economic Development	3
Federal Ministry of Agriculture	1
Federal Medical Insurance Fund	1
Federal Ministry of Culture	1
Federal Service for Control over Consumers' Goods (Rospotrebnadzor)	1
Total number of indicators	329

Bibliography

Barabashev, Alexei and Straussman, Jeffrey D. (2007), 'Public Service Reform in Russia, 1991-2006', *Public Administration Review,* 67 (3), 373-82.

Barber, Michael (2007), *Instruction to deliver : Tony Blair, public services and the challenge of achieving targets / Michael Barber* (London: Politico's) xiii, 418.

Downs, Anthony (1967), *Inside bureaucracy* (Boston: Little Brown and Co.) xv, 292.

Enikolopov, Ruben, et al. (2012), 'Field Experiment Estimate of Electoral Fraud in Russian Parliamentary Elections', *Proceedings of the National Academy of Sciences*.

Gueorguieva, V., et al. (2009), 'The Program Assessment Rating Tool and the Government Performance and Results Act Evaluating Conflicts and Disconnections', *American Review of Public Administration,* 39 (3), 225-45.

Hood, C. (2006), 'Gaming in targetworld: The targets approach to managing British public services', *Public Administration Review,* 66 (4), 515-21.

Jakobson, Lev (2001), 'Public management in Russia: changes and inertia', *International Public Management Journal,* 4 (1), 27-48.

Libman, A., Kozlov, V. , and Schultz, A. (2012), 'Roving Bandits in Action: Outside Option and Governmental Predation in Autocracies', (Frankfurt School - Working paper series №190).

Local Government Information Unit. (1994), *Performance indicators, a guide* (London: Local Government Information Unit).

Longenecker, C. and Ludwig, D. (1990), 'ETHICAL DILEMMAS IN PERFORMANCE-APPRAISAL REVISITED', *Journal of Business Ethics,* 9 (12), 961-69.

Pollitt, Christopher (2006), 'Unleashing Change: A Study of Organizational Renewal in Government', *Administrative Science Quarterly,* 51 (2), 302-04.

Sims, Jr Henry P., Gioia, Dennis A., and Longenecker, Clinton O. (1987), 'Behind the Mask: The Politics of Employee Appraisal', *Academy of Management Executive (08963789),* 1 (3), 183-93.

Smith, Peter (1995), 'On the unintended consequences of publishing performance data in the public sector', *International Journal of Public Administration,* 18 (2), 277 - 310.

Tullock, Gordon and Rowley, Charles Kershaw (2005), *Bureaucracy / Gordon Tullock ; edited and with an introduction by Charles K. Rowley* (Indianapolis, Ind.Lancaster: Liberty Fund;Gazelle Drake Academic) xvii, 437.

Verheijen, T. and Dobrolyubova, Y. (2007), 'Performance management in the Baltic States and Russia: success against the odds?', *International Review of Administrative Sciences,* 73, 205-15.

Якобсон, Л.Я. (2006), 'Бюджетная реформа: федерализм или управление по результатам', *Препринт WP8/2006/03, Государственное и муниципальное управление* (МОСКВА: ГУ ВШЭ).

Яковлев, А. (2010), 'Государственный капитализм, коррупция и эффективность госаппарата', *Общественные науки и современность,* 4, 18-25.