

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/78632>

Please be advised that this information was generated on 2019-02-19 and may be subject to change.



## Adapted Performance Measures and the Performance of Dutch Municipalities in Reducing Crime Rates

**Michiel S. de Vries**

Radboud University Nijmegen, The Netherlands

**Christiaan J. Lako**

Radboud University Nijmegen, The Netherlands

---

This article proposes to change performance measurement in such a way that it becomes possible to relate outcomes to the public policies involved, taking into account the complexity of the context. It involves a five step procedure, including quantitative and qualitative analysis. It is argued that this method can also be used as a search-light to detect best practices. This five step procedure is applied to the problem of crime rates in Dutch municipalities. The outcomes suggest that just looking at crime rates as such can distort conclusions about the effectiveness of municipalities. The remarkable finding for the Netherlands is that among those cities experiencing high crime rates there are nevertheless also cities having exemplary policies to combat crime.

**[Key Words: performance measurement, crime rate, municipalities]**

---

## I. INTRODUCTION

This paper proposes a change in the way of performance measurement in such a way that it can measure the effectiveness, sustainability and robustness of municipal public policies. Such policies aim at resilience, survival and adaptation and are specifically focused on controlling societal and organizational problems (Hood, 1991: 511). The fact that this is not measured until now is not just an academic problem, but also a problem in the development of new policies. If sayings are true that “what you measure is what you get” and “you measure what you are steering”, the absence of such measurement implies that new policies will not be or at least be less aimed at being effective, robust or sustainable. The sayings can, however, be disputed, because as this paper will argue, much policies are indeed aimed at sustainable and robust outcomes. However such policies don’t get the praise they deserve if their results are not measured or cannot be measured. If one keeps on measuring crime prevention by the number of crimes per capita – the judgment about metropolitan policies will stay negative, because no matter what the contents of their policies, their crime rate as measured will still exceed the crime rate in less densely populated villages.

For this problem this paper seeks an alternative. This paper sketches the outlines of a method to measure quantitatively the effectiveness and robustness of public policies. Before this proposal is made the paper concisely addresses the developments in measuring policy outcomes as seen in the recent literature and the attempts to improve these measurements. This part of the paper argues that these attempts are unsatisfactory and therefore it introduces the proposal. The proposal is sketched and applied to crime prevention policies in 443 municipalities in the Netherlands in 2006. The outcomes show that the method delivers on its promises. It is shown that this way of measuring policy outcomes is able to measure the effectiveness and robustness of policies and it can direct a researcher or policy maker towards best practices in the field, that is, to those

municipalities in which the policies are especially effective and robust. Therefore we suggest this method to be used more widely and also in other countries.

The resulting *main research question* is: Can the method for performance measurement to measure the effectiveness and robustness of municipal policies be further elaborated and validated in order to make it a valid and reliable method in various policy areas?

The *sub questions* to be answered are:

1. What developments are seen in the use of performance measurement at the local level?
2. What developments are seen in the scholarly literature on performance measurement and especially the measurement of policy outcomes?
3. How does the method as proposed fit into those developments and what are its merits?
4. To which degree is the method valid and reliable or does it need to be adapted given methodological criteria and the outcomes of the case-studies?
5. What does this imply for the possibilities of wider application of the method?

The paper is concluded by a request to react to this way of performance measurement. Of course, the authors could not oversee all the pros and cons. We did show our model before to municipal practitioners in the Netherlands and they were quite enthusiastic. However, this does not imply that the proposal is valid from a scholarly point of view.

## **II. THE BACKGROUND OF THE PROBLEM OF PERFORMANCE MEASUREMENT**

Many scholars in public administration have addressed the flaws in the present way of measuring performance of policies conducted in the public sector (Campbell, 1971; Dunn, 1981; Hood, 1991; De Vries & Lako, 2009; Bovaird & Loeffler, 2003, Kaplan & Norton, 1996, Halachmi & Bouckart 1996; Gomes, 2001; Sinclair en Zairi, 2000; Van Thiel & Leeuw, 2003; de Bruijn, 2001; Dijkstra, 2008; Ter Bogt & Van Helden, 1994; Hood, 1991).

This criticism does not remove the merits of performance measurement as is seen in the long tradition in its use. It is difficult to pinpoint the exact origins of performance measurement. Some emphasize the developments in the last decades under the influence of New Public Management (Hood, 1991). Others take the 1940s as the starting point (Mikesell, 1995), still others return to the scientific management approach of Frederick Taylor in 1920 with its application in the five year plans in Russia (Campbell, 1971) and Dunn mentions the early 19th century empiricism in England where even the amount of horseshit was measured (Dunn, 1981). We can go back in time even further to, for instance the Roman Empire, the Greek and Egyptian civilizations and to the antic Sumerian civilization where people were already accountable for their performance. Performance measurement has a long tradition and has become more popular recently. This is because it supposed merits.

Parts of the reasons are found in its functions. The advantages were summarized by Behn (2003). He mentions the possibilities it gives to evaluate, to motivate, to control, to reward, to budget, to learn and to improve (Behn, 2003: 588 ff.). Performance measurement can answer such questions such as: How well does the organization perform? How to ensure that people within the organization do what they are supposed to do? In which programs should one invest more money? How to motivate the staff to improve their performance? How to convince others that the organization does a good job? Which actions should be rewarded? (Pay for Performance) and can be seen as a success. What works, what doesn't? What could be done to improve performance? This implies that valid and

reliable performance measurement can be a crucial tool for management.

### **III. CRITIQUE ON PERFORMANCE MEASUREMENT**

However, as mentioned, the validity and reliability of performance measurement is under attack. Many scholars agree that it lacks and always has lacked validity which results in grave side-effects of attempts to measure performance. The critique addressing the validity and reliability of performance measurement, is seen in well-known sayings such as “What is measured is not relevant, and what is relevant is impossible to measure” or “what you measure is what you get” referring to the self fulfilling prophecies often contained in performance measurement thus disputing its validity. Such critique is sometimes formulated in such nice sayings, but often less beautiful connotations are also seen. An example of the latter is David Eddy who stated that 'Today's measures tend to be blunt, expensive, incomplete and distorting. And they can easily be inaccurate and misleading' (Eddy, 1998: 15-16).

Although the critique on the lacking validity is the most serious, also other problems of performance measurement are known. Some refer to the perverse impacts of performance measurement, the one-sidedness thereof, the absent relation with an organization's strategy or mission, and the lacking measures for long-term effectiveness as continuing problems (Kaplan & Norton, 1996; Boyle, 1989; Gomes, 2001). Sinclair & Zairi (2000) gave 10 reasons why performance measurement often does not work: 'Failure to define performance operationally, failure to relate performance to the process, failure to define the boundaries of the process, misunderstood or misused measures, failure to distinguish between control and improvement measures, measuring the wrong things, misunderstanding of information by managers, fear of distorting performance priorities, fear of exposing poor performance, perceived reduction in autonomy.'

(Sinclair & Zairi, 2000: 158)

Dutch scholars mentioned similar problems with performance measurement. The measures would create an artificial reality, sometimes contrary to the essence of reality; and they would have a negative impact on effectiveness (Van Thiel & Leeuw, 2003); They are often not thought-through; measure only what can be measured instead of measuring what is important, and they create perverse effects (de Bruijn, 2001).

#### **IV. THE SPECIFIC PROBLEM OF MEASURING OUTCOMES**

A specific problem of performance measurement addressed in this proposal is that such measures are directed mainly at measuring input, process and output, but cannot measure outcomes (Dijkstra, 2008). Ter Bogt & Van Helden (1994) even called the search for outcome measures illusionary. In the Netherlands Dijkstra argued in 2008 that this problem is especially detrimental for the public sector. We are still unable to measure those values that are of crucial importance for the public sector. He referred to the distinction made earlier by Christopher Hood in so-called sigma-type, theta-type and lambda-type values. Hood stated that especially the measurement of lambda values, in which sustainability and robustness of policies are crucial, are not (yet) developed. Especially the first type of Sigma values aimed at measuring cost-efficiency is well developed. But Theta values aimed at measuring ethics and integrity and Lambda values are underdeveloped.

According to Hood & Jackson this is a direct consequence of the emergence of New Public Management in the public sector (Hood, 1991: 4-5), because that trend emphasized 'Hands-on professional management,' explicit standards and measures of performance, greater emphasis on output controls; a shift to

disaggregation of units in the public sector, a shift to greater competition in the public sector; stress on private sector styles of management practice; and stress on greater discipline and parsimony in resource use. Also, the fact that an organization cannot be aimed at satisfying all three values simultaneously was important. To achieve Sigma values one sometimes needs to downsize whereas to achieve Lambda values, some redundancy, overlap and diversity might be crucial.

A related problem is that outcomes are often contingent, that is, dependent on the circumstances, while performance measures are mostly a-contextual. Behn kept his opinion on this problem brief: 'Abstract measures are worthless' (203, 598). Performance measurement is then simplified to the question: How much money was invested and which products were delivered in return? Recently Bovaird & Loeffler stated that: "there are two key areas in which measurement is required: improvements in public policy outcomes; and implementation by all stakeholders of a set of principles and processes by means of which appropriate public policies will be designed and put into practice." (Bovaird & Loeffler, 2003: 217) According to these scholars there are important trends visible in the process of performance measurement, but "While during the first generation of public sector reforms in the 1980s and 1990s a lot of lip service was paid to the need to measure outcomes, not much action really took place on this front"

## V. TERMINOLOGY

Outputs are originally defined the direct results of policies whereas outcomes are the effects thereof in terms of the goals. 'Outputs are defined as measures of activity or effort, while outcomes are the results of those services compared to their intended purposes.' (Kahn & Hildreth, 2004: 65) Outputs are the final products, whereas, outcomes are the more general objectives and results (Osborne & Gaebler, 1992). This is similar to the definition of outcomes of the American

Evaluation Forum, namely “a systematic way to assess the extent to which a program has achieved its intended results. The main questions addressed are: What has changed in the lives of individuals, families, organizations or the community as a result of this program? Has this program made a difference? How are the lives of program participants better as a result of the program?” (The Evaluation Forum, 2000: 9).

In this regard one of the main problems in measuring outcomes is how to determine whether the change in the problem is caused or induced by the policy that intended to diminish the problem. Often this question is difficult to answer, because there are many factors in the context that could also have been the cause of the visible change.

This brings us to a first criterion for outcome measurement. It should be able to identify the effect of a policy on the volume of the problem.

Second, robust policies are not just aimed at diminishing a policy problem, but also at diminishing the causal impact contextual factors might have on the problem. This refers to the robustness of policies. For instance, many policies are aimed at diminishing the probability that a child raised in a poor, a broken or otherwise deprived family will become a school dropout, becomes an adult showing criminal behavior, etc. Many policies are aimed at cutting such causal links. This brings us to the second criterion, namely that outcome measurement should be able to identify whether policies are able to bring problems under control given the contextual influences.

Therefore we define outcomes as the degree to which - taking influential contextual factors into account - problems are brought under control by a policy.

## **VI. PREVIOUS ATTEMPTS TO MEASURE OUTCOMES**

Several ‘solutions’ have been proposed before to measure policy outcomes.

Below it is argued that none of these is a solution for the problem of measuring the effectiveness, and robustness of policies. Sometimes one opts for qualitative measures. Performance measured for instance by surveys with questions such as: “Are you happy?” or “do you feel safe?” on the basis of site visit teams, audits based on self-evaluation reports and awards for best practices (Bovaird & Loeffler, 2003). Other solutions proposed were to measure outputs and to give a solid argumentation that the achievement of outcomes can be derived from these outputs. “We were successful in delivering these outputs and this implies - given this and that reasoning - that the outcomes will probably also be positive.”

The second solution found in the scholarly literature is that the concept of outcomes is redefined. First, some scholars put outcomes together with outputs under the common denominator of results (Halachmi, 1982; Wholey, 1983; Halachmi & Bouckaert, 1996: 4), being distinct from process measures. The solution implies that although one only measures outputs, the denominator of results, implies implicitly that also outcomes are measured. Another “solution” by redefinition is to see outcomes as outputs on a specific date further in time: ‘Outcomes are measured as the measures of effects of a program at some designated date’ (Propper et al., 2003). Other approaches include the definition of output as organizational goals and defining outcomes as the effects among the target group, or to define outputs as a phenomenon measured statically and outcomes as the change therein (Gleason & Barnum, 1982) or to see outputs as measured by single indicators and outcomes as the result of a composite index.

A third type of approach is to shift the concepts of input, process, output and outcomes. The process is then seen as output (for instance the number of lecturing hours, meetings, training sessions, or police-surveillance) and the output as outcomes (the number of diplomas, the number of fines, captured criminals).

Last but not least, an elegant solution is to see outcomes as those measures about which stakeholders agree that they are outcomes.

The problem with all these “solutions” is, however, that they do not solve the

problem how to measure outcomes as defined before. Therefore, renewed research into the possibilities of new methods for performance measurement is severely needed.

## **VII. A MIXED METHOD FOR PERFORMANCE MEASUREMENT AND THE IDENTIFICATION OF BEST PRACTICES**

Taking our own definition of robust and effective policies serious, that is, the degree to which - taking influential contextual factors into account - problems are brought under control by a policy, one should measure the extent to which policies are able to do so and do take important contextual factors into account. Below a procedure to measure policy outcomes involving five steps is proposed.

1. The construction of a causal model which relates the policy problem to its causes found in the context where the problem occurs (In order to make performance measurement contextual).
2. The testing of this model using statistical data and regression analysis over municipalities (because the number of municipalities allows statistical research and the wide availability of data on municipalities). This method has become feasible for measuring policy outcomes in municipalities in the Netherlands with the construction of several public databases on Dutch municipal policies. These can be found at websites such as [www.waarstaatjegemeente.nl](http://www.waarstaatjegemeente.nl), [www.overheidsmonitor.nl](http://www.overheidsmonitor.nl), [www.CBS.nl](http://www.CBS.nl).)
3. The additional incorporation of policy outputs for every municipality into the regression model (In order to measure the added value (effectiveness) of the municipal policy controlling for the contextual factors. In this way the effectiveness of municipal policies can be measured.)
4. Analyzing whether the policy is able to cut the relation between the problem

and the causal factors in the context (i.e. measuring the difference between slopes with and without taking the policy into account in order to measure the robustness of the policy, that is, making social problems less vulnerable for contextual changes).

5. Determining best practices by:

- a. Measuring the residues (that is the difference between the expected value of the problem and the real value of the problem in all municipalities). The values of the residues for each municipality are used as a search light for finding those municipalities in which it is likely that there exists a best practice.
- b. Doing a content analysis on the policies of those municipalities in which the residue is largest (in order to analyze what it is that makes the problem in these municipalities so much smaller than expected.)

As seen this five-step method proposes a mixed method. Partly statistical and quantitative and partly qualitative using the case study approach.

*sub 1. Towards a context related measurement of the problem*

The first step in the model is to generate an indicator of the size of the policy problem, taking into account contextual factors. For this purpose we compare all Dutch municipalities for their crime rate. (cf. [www.waarstaatjegemeente.nl](http://www.waarstaatjegemeente.nl), [www.overheidsmonitor.nl](http://www.overheidsmonitor.nl), [www.CBS.nl](http://www.CBS.nl)). A basic model is produced in which one can include generally available data such as population, demographics, economic growth and welfare, or the percentage of all houses being rented, and unemployment, and one can add specific variables of which it is known from scientific research that they impact on crime rates. The only restriction in this step is that only factors are included which are outside the influence of the policy maker. The basic factors are together with the specific factors regressed on the size of the policy problem of which one wants to know whether the municipality

conducts robust and efficient policies. Based on the regression analysis one is able to determine the expected values of the size of the policy problem given the features of the context.

Comparing the real size of the problem with the expected size, enables one to judge whether a municipality is doing rather good or bad in this area (residue analysis).

*sub 2. Measuring the effectiveness of policies*

Characteristics of the municipal policy are added to the model in the second step. For instance, one can take the amount of money invested as given in the budget, or the number of people that are paid to solve the problem or whichever variable available. The goal is to analyze whether such policies are of significant effect in reducing the size of the problem given the features of the context. This is in our point of view indicative for the effectiveness (outcomes) of the policy in the specific area and indicative for the question whether the effects are indeed caused by the policy. By determining the ratio between the expected and real size of the problem one can even make statements about the value for money created by such policies. Finally, by conducting longitudinal research one can judge whether the effectiveness of such municipal policies increases or diminishes over time.

*sub 3. Measuring the robustness of policies*

The analysis in step 2 can show that the regression coefficients of factors from the base model diminish after the adding of policy variables. This is seen in the difference between the coefficients in the base model and the same coefficients in the model after one adds the policy factors into the model. When such interaction effects are visible, one can conclude that the problem is made less context dependent, which is indicative for the degree of robustness.

*sub 4. Measuring the sustainability of policies*

If data are available of several years one can build so-called lags into the model to judge whether the policy instruments used are effective in diminishing the problem on the long term, determine after how many years they lose their impact (the longer the more sustainable) and determine whether they are only effective in the year they are implemented (No sustainability).

*sub 5. Determining best practices.*

The last step is included to Judge whether in a specific municipality the instruments applied are especially effective, robust or sustainable and thus whether it is probable that a best practice exists. This can be done in three ways:

- First one can analyze for every municipality separately whether the size of the policy problem is smaller or larger than might be expected on the basis of the base model. The difference (by residue analysis) is indicative for the answer to the question whether a municipality is doing better or worse in this area than others, irrespective of their policies.
- Secondly one can analyze whether there exist municipalities that have the problem much better under control than expected. If the real and expected size of the problem are far apart in favor of the expected value, this could imply that a best practice exists of which the features are not incorporated in the model. The residue analysis can steer the search for such practices.
- Thirdly, one can analyze on the basis of the extended model whether a policy in a specific municipality is relatively better than that in other municipalities. This is possible by adding interaction effects in the model.

## **VIII. A CASE STUDY INTO LOCAL CRIME PREVENTION POLICIES IN THE NETHERLANDS**

The variance in the number of crimes in municipalities in the Netherlands

results in interesting rankings (see a/o Politie en Wetenschap, 2008). Based on such figures as collected on websites such as [www.watdoetjegemeente.nl](http://www.watdoetjegemeente.nl) one is able to produce the following ranking for 2006.

<Table 1> Top 10 municipalities on number of crimes per 1000 capita in 2006

Most unsafe municipalities			Most safe municipalities		
1)	Utrecht	147.5	1)	Ferwerderadiel	18.2
2)	Amsterdam	143.6	2)	Littenseradiel	19.1
3)	Eindhoven	125.5	3)	Leeuwarderadeel	19.6
4)	Maastricht	121.1	4)	Liesveld	19.8
5)	Ouder-Amstel	120.1	5)	Opsterland	21.0
6)	Roermond	119.0	6)	Ten Boer	21.4
7)	Nijmegen	119.0	7)	Graafstroom	21.9
8)	Rotterdam	118.0	8)	Menaldumadeel	22.0
9)	Hertogenbosch	116.1	9)	het Bildt	22.2
10)	Arnhem	112.6	10)	derkerk	23.2

*Note 1.* These are mostly large municipalities.

*Note 2.* These cities are hardly known outside the Netherlands. This is because they belong to the smallest municipalities in the country.

The number of crimes per capita over all Dutch municipalities is a fairly normal distributed variable, somewhat skew with a mean of 55.3 and a median value of 51.9 crimes per 1000 inhabitants in which 95% of all municipalities score between 53 and 57.5. From recent scholarly research it is known that the degree of urbanization (Faris & Dunham, 1965; Traub & Little, 1999), poverty (Traub & Little, 1999), cultural diversity (De Beer & Schuyt, 2004) influence the crime rates. These contextual factors can hardly be tackled by public policies on security.

The values for these factors is known in the Netherlands for each municipality ([www.cbs.nl](http://www.cbs.nl); [www.waarstaatjegemeente.nl](http://www.waarstaatjegemeente.nl), [www.watdoetjegemeente.nl](http://www.watdoetjegemeente.nl)) through indicators as the number of inhabitants, the percentage of children growing up in deprived families, the number of families living on social security, and the number of immigrants. The regression of these factors on the endogenous

variable, results in an explanatory model that explains 41% of the variance in the crime rates at the municipal level ( $R^2 = .41$ ). The percentage of children living in deprived families is the most important explanatory factor, with a “slope” of 3.43 ( $\beta = .49$ ,  $p < .000$ ). Table 2 presents the outcome of this analysis in the upper model 1. All independent factors have the expected and significant effect on crime rates ( $N = 442$  municipalities,  $p < .05$ ).

<Table 2> Outcomes of the regression model on crimes in Dutch municipalities in 2006

		Unstandardized Coefficients	Standardized Coefficients		t	Sig. <sup>c</sup>
		B <sup>a</sup>	Std. Error	Beta <sup>b</sup>		
Model 1	(Constant)	37,753	1,400		26,968	,000
	Children in special needs families population 2006	3,438	,300	,489	11,459	,000
	Percentage of immigrants	,058	,012	,199	4,695	,000
		,370	,179	,080	2,070	,039
R <sup>2</sup> Model 1 = 0,40						
Model 2	(Constant)	67,742	2,889		23,451	,000
	Children in special needs families population 2006	4,046	,269	,576	15,060	,000
	Percentage of immigrants	,053	,011	,182	4,879	,000
	Percentage of crimes solved	,314	,157	,068	1,997	,046
	-1,412	,123	-,376	-11,472	,000	
R <sup>2</sup> model 2 = 0,56						

a. The unstandardized coefficient gives the effect of the factor in question on crimes per 1000 inhabitants

b. These are the standardized regression coefficients

c. This is the level of significance of the regression coefficients.

Note1. Dependent variable is crimes per 1000

In model 2 we added the factor “risk of being caught” that is the number of solved crimes. The explained variance,  $R^2$ , increases to 55%. This implies an

increase of the explanatory power by this one factor of 14%.

### 1. Effective but not robust policies

It is important to notice that the risk of being caught does have a direct effect on diminishing crime rate; however, it hardly diminishes the influence of the contextual factors on crime rates. The impact of urbanization, poverty and cultural diversity hardly diminishes. Hence, policies aimed at an increased risk of being caught are unable to break through the causal relation between contextual factors and the size of the problem. Hence, solving more crimes is in itself unable to control the impact of the context on the problem.

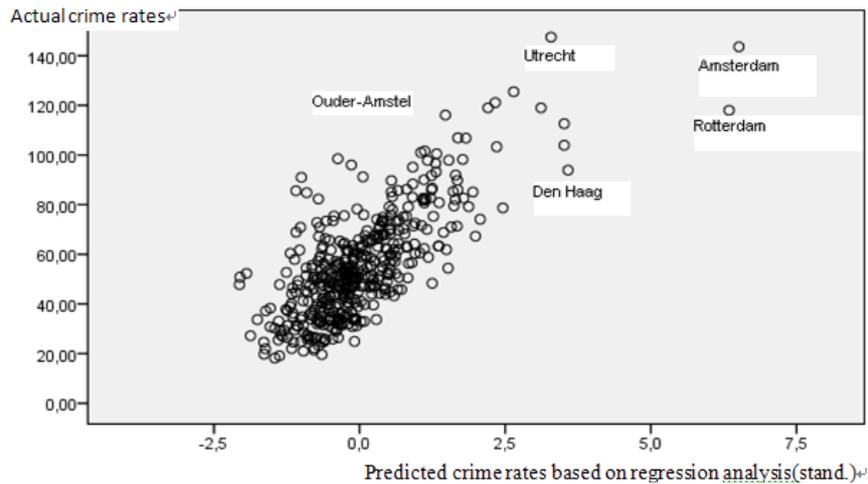
### 2. Searching for a best practice: The Rotterdam 'coercion and urge' policies

The question arises whether there are municipalities that in all likelihood do conduct robust policies and where one sees a much lower crime rate than expected. Based on the regression analysis the expected value of the crime rate is related to the real crime rate. These are presented in the scatter plot given below. For some interesting municipalities the name is added to the scatter plot.

The scatter plot shows a fine relation between predicted and real values of crime rates. However, there are some municipalities in which the difference between the two is rather large. Especially the position of the city of Rotterdam is outstanding. The actual crime rate in this city is almost 30% lower than one would expect on the basis of urbanization density, poverty, cultural diversity and the risk of being caught. Normally one could see Rotterdam as an outlier and remove it from the analysis. However, given our five step model, we assume that there is a high probability that Rotterdam conducts a best practice in fighting crime. This is not necessarily the case, but we assume the analysis provides a search light, in this

case pointing to Rotterdam. If there are specific policies conducted which are especially effective in reducing the crime rate it is likely that these are to be found in the city of Rotterdam.

<Figure 1> Actual crime rates in 2006 and predicted values based on the regression analysis



Did Rotterdam indeed conduct something like a best practice trying to cut the link between context and problem? An elaborate policy analysis is needed to confirm this, but has to be disposed of because it would require a separate paper. Nonetheless the security policies in Rotterdam have changed significantly during the last couple of years. Dominant are phrases such as ‘zero-tolerance’, a tough approach against what are even in official policy documents called ‘shitheads’, a tight policy on social housing in order to keep the needy out of certain neighborhoods (COS, 2008a) and a social security policy in which the principle “work first” is central (COS, 2008b). People, of whom it is expected that they abuse social security, or of whom it is known they cause nuisance, are forced to work at the garbage collection service. It is the so-called ‘coercion and urge’

approach. In 2003 the city developed the plan to get the 700 drugs users causing most nuisances off the streets. Before 2005 all these 700 people should either be prosecuted or enter a social assistance program. The new policies would not only be more effective, but also more obligatory (cit. AD 2 august 2006). In 2006 alone 93 administrative sanctions were in order against what the Dutch call ‘coffee shops’ and against shops and persons that caused nuisances. Licenses were revoked, shops were closed and legal penalties were given (Municipality of Rotterdam, 2008). Whether this particular policy is robust - in our definition - is unsure. It requires a lot more qualitative and quantitative research to draw the conclusion that the Rotterdam policies are indeed able to break down the relation between the adverse contextual factors and the crime rate. These aspects of security policies do distinguish Rotterdam from other municipalities.

This section illustrated the five step procedure to identify best practices. It argued that such best practices are not found by simple ratio-analyses, but that taking contextual and policy factors into account a rather different picture emerges in which a municipality (Rotterdam) in which the crime rate is highest, can still conduct best practices in the security area. The example illustrates that the procedure seems to be valid and reliable and can be an adequate tool to search for best practices. Below the merits of the procedures are further discussed.

## **IX. DISCUSSION ON THE MERITS OF THE METHOD**

The last question this paper addresses is how this way to measure performance relates to other methods and whether it is an improvement compared to those other methods. Let’s be clear about it, the method using residues as a way to search for best practices is known. Especially in the area of public health (Hammond, 1986; Hatry & Fisk, 1992; Nyhan & Martin, 1999). Smith (1999) has given a discussion of the merits of multivariate regression analysis that allows for

the development of adapted performance measures, when a large number of cases are involved. However, the method is hardly used in public administration. The method was criticized because the standard - the regression line - is variable instead of constant, because the standard is based on average performance instead of optimal performance, because the estimates based on quantitative analyses are often biased, and because only one dependent variable at the time can be taken into account (Kopsczynski & Lombardo, 1999: 131; Maddala, 1983; De Lancer, 1996: 19; Nyhan & Martin, 1999). In comparing businesses with operating in similar circumstances, it might be sensible to take optimal performance as the standard. However, in comparative public administration there are sound arguments to make a standard variable as seen in the example given above. Crime rates are expected to be higher in large cities than in small villages. And in this case, with varying circumstances, fixed standards result in awkward outcomes as was seen in table 1. The smallest villages have the smallest crime rates. This does not imply they also have the best security policies. Many of them don't even have nor need security policies. It is only because the crime rates are not related to the context or to the policies that this seems to be the case. Furthermore, comparing to average or best performance is relative. It is not the absolute distance to the standard that determines performance. The best actual scores compared to the relative scores determine this.

### 1. Validity

The biggest advantage of the five step plan is according to us that it produces outcomes that are context related and are related to the policies. It is a valid way to measure the core of policy making, that is, to bring a problem under control and to make sure that a robust and effective answer is given to problems irrespective of the adversity of the context. Often, the real causes of a policy problem cannot be resolved - compare lightning which can bring about fire. However, policies

can be seen as a lightning rod, in which policy makers aim to break down the causal relation between lighting and the starting of fire. In policies, often it cannot be prevented that children grow up in underprivileged families and neighborhoods. Policies can aim to prevent such children from what the Rotterdam policy makers call shitheads. A policy maker often wants to do more than just extinguish a fire. To prevent it from happening is as important. The breakdown of the causal relations between background and problem can be measured through the five step procedure by comparing the slopes on the problem without taking the policies into account and when taking the policies into account.

Other criteria to determine to judge the method are found in the usefulness, feasibility, possibilities of abuse, accuracy and impact.

## 2. Usefulness

Performance measurement intends to provide information that is useful. The outcomes of the proposed measurement give such information. It gives information about the size of problems given the specific situation, whether a policy diminishes the problem and whether and to which extent the policy brings the problem under control, that is, less dependent on contextual factors which are out of control. Furthermore, this way of measurement gives a more meaningful direction to the search for best practices.

These merits have a counterpart in that this kind of measurement is more complicated and less user friendly. The terminology might deter users. However, the outcomes don't have to be articulated in statistical terms. They can also be phrased in terms like: "given the characteristics of our municipality we are doing exceptionally well in the area of..." or "given that the real size of the problem is much lower than expected in such a municipality..." or "Our policies have resulted in that we are better able to control the problem, or "Even when an extraordinary situation arises, we will be able to limit the consequences thereof".

### 3. Feasibility

Is it realistic to expect that such performance measurement will be widely used? This is dependent on the availability of data, knowledge of statistical analysis, the ability to apply such techniques and the insight to give a proper interpretation of the outcomes, as well as the time available to conduct such analyses. As far as the availability is concerned, we encountered little problems. However, this might be different in other countries. In the Netherlands there are easily available monitoring systems giving data at the local level. However, adequate knowledge of regression analysis is not widely available. Only in those municipalities that have statistical departments and in which employees did finish a higher education in which statistics is part of the curriculum this will be the case. As far as time is concerned, it might be said that an experienced researcher can conduct such an analysis in one day. Of course the first time it will be more time consuming, but routine will make such analysis easier to conduct.

### 4. Manipulation possibilities

Every method of performance measurement has the capability of being abused. This way of measurement is no different. One could expect that the increased complexity will result in a larger probability of abuse. First of all, one can manipulate the outcomes within the base model. The operationalization of the variables and the factors included and excluded, influence the outcomes and the ranking. However, the method is only manageable if data are used that are comparable over different units. In our case the data have to be measured in the same way in every municipality. This diminishes the probability of bias. The complexity can also be seen as advantageous in this respect. Perverse consequences of performance measurement are made more difficult, because the

effects thereof cannot be estimated beforehand. This makes such behavior less expedient.

## 5. Precision

The precision of performance measures is dependent first and foremost on the quality of the data and the quality of the analysis. It is necessary to build an adequate base model that contains the most important contextual factors. If that base model is adequate and the data are valid and reliable the model will deliver on its promises. These conditions are not always met. Especially when one does not conduct an analysis on municipal policies, but instead conducts an international comparison on for instance literacy rates, fundamental problems do arise (De Vries, 2009b). Hence, the usefulness of the analysis does depend on the quality of the data and the analysis model.

## 6. Impact of the measurement

This paper investigated primarily the possibility to use quantitative measures of performance. Irrespective how scholars feel about the merits of quantitative versus qualitative analysis, in the practice of public administration the power of numbers as such should not be underestimated. Conclusions based on qualitative research often have less impact than conclusions based on statistical analysis. One can regret this, but it seems to be a fact of life. Especially in benchmarking statistical analysis has huge impact. A municipality will react when the conclusion based on a broad statistical comparison is that their policies are less effective, robust and sustainable than the policies in a neighbouring municipality. This is why statistical comparison is one of the steps in this analysis. However, it is not the only step. It has to be followed by an investigation into qualitative differences. This is suggested in the last step of our model. Only the combination will result in

learning effects.

## X. CONCLUSIONS

This paper proposed a model on performance measurement that takes contextual conditions into account and relates the outcomes to actual policies. The goal is to remove the problems inherent in ratio analysis as indicative for performance. Those problems can be summarized as follows. Performance measurement is too often a-contextual, measures inputs and outputs instead of outcomes and is difficult to relate to implemented policies.

This paper proposed a mixed method consisting of a five step model to address these problems and to enable a more valid measurement of performance.

In the first step it is investigated whether a specific problem in municipalities is more or less severe than could be expected on the basis of characteristics of those municipalities. The second step involves the analysis whether municipal policies are effective in reducing the problem. Subsequently the third step analyses whether the municipal policies are able to control the problem, that is, break down the relation between the contextual factors and the problem. The fourth step consists of measuring whether the municipal policies are able to produce long-term effects and the fifth step is to conduct a content analysis on the policies of those municipalities where the problem is all in all much smaller than expected in order to judge whether there is indeed something like a best practice in that municipality.

This model was applied to Dutch municipal policies on combating crime. This resulted in the striking outcome that the municipality which is in the top 5 of municipalities regarding the severity of the problem, at the same times seems to conduct the best policies fighting crime.

Finally the paper addressed the merits of the model. It argued that despite

earlier criticism on residue analysis, the advantages of using such analysis outweigh the drawbacks. That this conclusion was drawn could be expected. Otherwise we would not have proposed the model. However, we would like to invite readers to react. As already said in the introduction, we could not take all possible advantages and disadvantages into account and it is probable we missed something. Therefore the reader is invited to think about the question whether the model makes sense, where it should be adapted and where it goes wrong.

## REFERENCES

- Andrews, C. W. (2008). Best practices in local government, In M.S. de Vries, PS Reddy & Shamsul Haque (eds.), *Improving Local Government: Outcomes of Comparative Research*. New York: Palgrave. MacMillan. Basingstoke.
- Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, 63(5): 586 - 606.
- Bies, A. (2004). Scanning for best practices, In Michael J. Austin & Karen M. Hopkins (eds.), *Supervision as Collaboration in the Human Services: Building a Learning Culture*. Thousand Oaks: Sage Publications, 261-271.
- Bogan, C. E. & English, M. J. (1996). Benchmarking for best practices, In R. L. Craig (ed.), *The ASTD Training and Development Handbook: A Guide to Human Resource Development*. McGraw-Hill Professional, 394-412
- Bogt, H. J. ter, G. J. van Helden. (1994). Kwaliteit van prestatiemeting bij gemeentelijke overheden, *Bestuurskunde*, 3(1): 2-12.
- Bovaird, T. & Löffler, E. (2003). Evaluating the quality of public governance: Indicators, models and methodologies. *International Review of Administrative Sciences*, 69(3): 313-328.
- Boyle, R. (1989). *Managing Public Sector Performance*. Dublin: Institute of Public Administration.
- Bruijn, H. de. (2001). *Prestatiemeting in de publieke sector: tussen professie en*

*verantwoording*. Lemma.

- Campbell, D. T. & Stanley, D. C. (1966). *Experimental and Quasi Experimental Designs for Research*. Chicago: Rand McNally.
- Chandna, R. C. (1986). *A Geography of Population: Concepts, Determinants and Pattern*. New Delhi, India: Kalyani.
- Coers, M. & Gardner, Chris. (2001). Benchmarking: A Guide for your journey to best-practice processes, *APQC*, 5.
- COS Centrum voor Onderzoek en Statistiek (2008a), *Evaluatie Wet Herhuisvesting Rotterdam*. Gemeente Rotterdam.
- \_\_\_\_\_. (2008b), *Werk Direct*, Gemeente Rotterdam.
- Dawson R., Hancock, D. R. & Algozzine, R. (2006). *Doing Case Study Research: A Practical Guide for Beginning Researchers*. Teachers College, Columbia University
- De Beer P. T. & Schuyt (red), C. J. M. (2004). *Bijdragen aan waarden en normen, WRR*. Amsterdam: Amsterdam UP.
- De Castell, S., Luke, A. & Egan, K. (eds), (1986). *Literacy, Society, and Schooling: A Reader*. Cambridge: Cambridge University Press.
- De Vries, M. S. & Lako, C. J. (2009a). *Prestatiemeting voor effectief en robuust beleid. Bestuurswetenschappen*, forthcoming.
- De Vries, M. S. & Lako, C. J. (2009b). Performance measurement and the search for best practices. Paper at the IIAS conference, Helsinki.
- Dijkstra, G. (2008). Outputmeting van overheidspersoneel: een simpele prestatie?, *Bestuurswetenschappen*. 1: 54-68.
- Dunn, W. N. (1981). *Public Policy Analysis: An Introduction*. Englewood Cliffs: Prentice Hall.
- Eddy, D. M. (1998). Performance measurement: Problems and solutions. *Health Affairs*, 17(4): 7-25.
- Epstein, P. D. (1996). Redeeming the promise of performance measurement issues and obstacles in the United States, In A. Halachmi & G. Bouckart, (Eds.), *Organizational Performance and Measurement in the Public Sector*. London: Quorum Books.

- Faris, R. E. L. & Dunham, H. W. (1965). *Mental Disorders in Urban Areas*. Chicago: University of Chicago Press.
- Gemeente Rotterdam. (2008). *Jaarverslag Bestuurlijke Handhaving 2007*. Rotterdam.
- Gerring, J. (2007). *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- Gleason, J. M. & Barnum, D. T. (1982). Toward valid measures of public sector productivity: Performance measures in urban transit. *Management Science*, 28(4): 379-86.
- Gomes, M. B. (2001). *Performance Audit Management: a Public Management Policy Analysis about Supreme Audit Institutions Role*. Caracas: XV concurso de Ensayos del Clad.
- Halachmi, A. & Bouckeart, G. (eds.), (1996). *Organizational Performance and Measurement in the Public Sector*. London: Quorum Books.
- Halachmi, A. (1982). Purpose and perspective in evaluation: A Case Study. *Administrative Science Review*, 12(2): 1-16.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1): 3-19.
- Hood, C. (2007). Public service management by numbers: Why does it vary? Where has it come from? What are the gaps and the puzzles? *Public Money & Management*, 27(2): 95 - 102.
- Kahn, A. & Hildreth, W. B. (2004). *Financial Management Theory in the Public Sector*. New York: Praeger publishers.
- Kaplan, R. S. & Norton, D. P. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Cambridge: Harvard UP.
- Keehley, P. & MacBride, S. A. (2002). Can benchmarking for best practices work for government? In M .G. Hartmann, *Fundamental Concepts of Quality Improvement*, ASQ Quality press, 313-324.
- Kopczynski, M. & Lombardo, M. (1999). Comparative performance measurement: Insights and lessons learned from a consortium effort. *Public Administration Review*, 59(2): 124-134.
- Lerner, D. (1959). *The Passing of Traditional Society*. Glencoe: The Free Press.

- Löffler, E. (1999). The value of public management evaluations from an international perspective. Paper presented at the international public management network workshop. Siena.
- Lonti, Z. & Gregory, R. (2007). Accountability or countability? Performance measurement in the New Zealand public service, (1992 - 2002). *Australian Journal of Public Administration*, 66(4): 468-484.
- Mikesell, J. (1994). *Fiscal Administration: Analysis and Applications for the Public Sector*. Belmont: Wadsworth Publishing Company.
- Nyhan, R. C. & Martin, Lawrence L. (1999). Performance measurement: A primer on data envelopment analysis. *Public Productivity & Management Review*, 22(3): 348-364.
- Osborne, D. & Gaebler, T. (1992). *Reinventing Government: How the Entrepreneurial Spirit Is Transforming Government*. New York: Penguin Books.
- Propper, C. & Wilson, D. (2003). The use and usefulness of performance measures in the public sector. *Oxford Review of Economic Policy*, 19: 250-267.
- Sinclair, D. & Zairi, M. (2000). Performance measurement: A critical analysis of the literature with respect to total quality management, *International Journal of Management Reviews*, 2(2): 145 - 168.
- Stake, R. (1995). *The Art of Case Study Research*. Thousand Oaks: Sage publications.
- The Evaluation Forum (2000). *The Outcomes for Success*. Retrieved at <http://www.evaluationforum.com/>
- Thiel, S. van & Leeuw, F. L. (2003). De prestatieparadox in de publieke sector, *Beleidswetenschap*, 17(2): 123-144.
- Traub, S. H. & Little, C. B. (1999). *Theories of Deviance*. Itasca, Illinois: F.E. Peacock Publishers.
- Wholey, J. S. (1983). *Evaluation and Effective Management*. Boston: Little and Brown.
- Wildavski, A. (1979). *Speaking Truth to Power. The Art and Craft of Policy Analysis*. Boston: Little, Brown.
- Yembuu, B. & Munkh-Erdene, K. (2005). *EFAA Global Monitoring Report Literacy: Country Study Mongolia*. Ulanbatar: Mongolian National Commission for UNESCO.

Yin, R. K. (2003). *Case Study Research: Design and Methods*. Thousand Oaks: Sage publications.

**Michiel S de Vries** is full professor and chair of the department of Public Administration of the Radboud University Nijmegen, the Netherlands. He has written numerous articles on comparative public administration, policy change and international technical assistance. He is member of the editorial board of international journals such as *International Review of Administrative Sciences* and *Acta Politica*. He is the chair of the working group on Local Governance and Development of IASIA, the chair of the working group on Public Administration Reform of NISPACee, and vice president of IASIA.

**Christiaan J. Lako** is an assistant professor at the department of Public Administration of the Radboud University Nijmegen, the Netherlands. He received a master in sociology from the Utrecht University in 1979 and worked since then at the Wageningen University, the Erasmus University Rotterdam, and at the Nijmegen department of Public Administration. He published in national and international journals about public management, performance measurement, co-payments, market competition and evaluation of health systems



## 범죄율 감소에 대한 네덜란드시 정부의 성과와 성과측정방법

Michiel S. de Vries

Radboud University Nijmegen, The Netherlands

Christiaan J. Lako

Radboud University Nijmegen, The Netherlands

이 논문은 상황의 복잡성을 고려하여 공공정책의 결과와 연계하는 방식으로 성과 측정하는 방식을 변화시키는 방법을 제안하고 있다. 이 과정은 다섯 가지의 단계로 구성되며 양적인 분석과 질적인 분석을 모두 포함한다. 저자들은 이 방법이 우수사례를 찾아내는 좋은 탐색기법이 될 수 있다고 제안한다. 이 5단계 분석방법이 네덜란드 지방정부의 범죄율과 연관된 문제에 적용되었다. 분석결과에 따르면 단지 범죄율이라는 결과만을 검토하는 결과지표는 지방정부의 효과성에 대한 분석결과를 왜곡할 수 있다는 점을 보여준다. 네덜란드에서의 이와 같은 연구는 높은 범죄율이 나타나고 있는 도시에서 범죄에 매우 잘 대응하고 있는 정부를 발굴하고 모범이 되는 정책을 찾을 수 있는 가능성을 부여하는 것이다.

[주제어: 성과측정, 범죄율, 지방정부]

논문접수일: 2009년 4월 15일 / 게재확정일: 2009년 5월 29일

제1저자: Michiel S. de Vries

소속 및 직위: Professor, Department of Public Administration, Radboud Universiteit Nijmegen

주소: Radboud University Nijmegen, dept. of Public Administration, Thomas van Aquinostraat 5,  
P.O. Box 9108, 6500 HK Nijmegen, The Netherlands

전화번호: +31-24-3615691

이메일: m.devries@fm.ru.nl

제2저자: Christiaan J. Lako

소속 및 직위: Center for Policy Analysis, Institute for Management Research

주소: Radboud University Nijmegen, dept. of Public Administration, Thomas van Aquinostraat 5,  
P.O. Box 9108, 6500 HK Nijmegen, The Netherlands

전화번호: +31-24-3613077

이메일: c.lako@fm.ru.nl

