# PERFORMANCE MEASUREMENT

## Its Legitimacy and Its Efficacy

*Anne Ruth Mackor*[1]

**Abstract**

In this paper it is argued that performance measurement deserves attention, not just of economists and scholars in business administration, but also of legal scholars. Section one explicates the notion of performance norms and explains how they relate to so-called 'duty of care' norms. Section 2 focuses on the history of performance measurement. It explains why the government, from the 1980-ies on, has begun to apply performance measurement to its own and other semi-public organisations. Section 3 deals with the question whether government enforced performance measurement is a legitimate way to regulate the behaviour of people and organisations. Section 4 discusses some doubts about the efficacy of performance measurement. Section 5 concludes that both the legitimacy and the efficacy of performance measurement deserve further study.

**Keywords**: performance measurement, legitimacy, efficacy

## 1 Introduction

At first sight, there cannot be any misunderstanding about the question what performance measurement is. Performance measurement is, quite simply, the measuring of performances. Performance norms enable us to do so. More specifically, they tell us *what* performance should be accomplished and the also contain *indicators* that enable us to measure whether the performance has been realised.

### 1.1 'Duty of care' norms and performance norms

The role of performance norms has become increasingly important in governmental regulation of human behaviour. This is a consequence of the fact that the legislator[2] nowadays often makes use of so-called duty of care norms. Duty of care norms differ in several important respects from classical legal rules. Classical legal rules are norms that forbid or prescribe

1 University of Groningen, Faculty of Law, Department Theory of Law, P.O.Box 716 9700 AS Groningen The Netherlands, A.R.Mackor@rug.nl

2 Unless stated otherwise, I refer to the Dutch government and the Dutch legislator.

certain types of *behaviour*, such as norms that forbid to drive a car when under the influence of alcohol, and that oblige driver and passengers to wear a safety belt.
A duty of care norm on the other hand, does not prescribe or forbid any specific kind of behaviour. Instead such a norm tells the addressee to realise some, more or less specified and more or less concrete, *goal*. The norm does not tell the addressee, however, through which behaviour the goal is to be accomplished. It is, in principle, up to the discretion of the addressee to choose his own means to fulfil his duty. For example, had the legislator promulgated a duty of care norm for car drivers instead of a behavioural rule, the rule would not forbid driving cars after drinking alcohol. Rather, the rule would consist of a duty to take care of the traffic safety, for that is the main goal behind the behavioural rules concerning alcohol and safety belts.

Duty of care norms seem to have one obvious advantage over behavioural norms, viz. that the addressee himself can choose the most effective and efficient means to realise the goal (provided of course that the means are legitimate). The legislator has been aware, however, that a duty of care norm, on its own, does not give enough direction to the addressee. Therefore complementary norms are needed. What is of interest here, is the fact that that it is not the legislator who promulgates these complementary norms. Instead, the legislator instructs the norm addressees to make and apply these complementary norms more or less 'on their own'.

Part of these complementary norms are performance norms. Performance norms are therefore important means, not only for the addressees, but also for the government, to determine whether addressees have fulfilled their duty of care.

This may sound rather abstract so let me introduce an example. Article 2 of the Dutch Act on the Quality of Health Care Institutions (Kwaliteitswet Zorginstellingen) is a duty of care norm which instructs the provider of care to offer *sound care* ('verantwoorde zorg'). Article 2 explicates sound care as 'care of a good quality that is at a minimum effective, efficient, patient oriented and tuned to the realistic needs of the patient'. This is all the Act tells the addressees about the goal they have to achieve. Nevertheless, the legislator does not leave it completely to the discretion of the health care institutions to determine the way in which they choose to fulfil their duty of care. Article 3 and especially 4 and 5 give further instructions.
Article 3 tells the care provider to *organise* the care in such a manner that the result is, or at least reasonably must be, sound care. Article 4 explicates *how* care should be organised in order to result in 'sound care'. Article 4 therefore is, at least for the purpose of this paper, the crucial article of the Quality Act.

Without mentioning the terms 'performance measurement', 'performance norms' or 'performance indicators', the article nevertheless obliges care providers to measure their own performances.

Article 4 instructs care providers:
- 'to systematically *collect and register* facts about the quality of care'
- 'to systematically *check* whether the organisation of care as meant in article 3 does *in fact* result in sound care',
- 'if necessary, *to change* the organisation of care.'

Article 5, finally, demands of the care providers to give account of the quality of the care that they have been offering by publishing a *yearly public report*. As article 4 and 5 make clear, performance measurement serves more than one goal. The primary aim is to systematically *register* the quality of care that has been offered. The aim of registration is to enable others (the government, but also e.g. patients and consumer organisations) to *assess* whether sound care has been offered. A further aim is the *improvement* of the quality of care. Finally, performance measurement enables care providers to *give account* of the quality of care they have offered.

Moreover, performance measurement has several other goals that are not mentioned in the Quality Act. Performance measurement is also, among others, the foundation of subsidisation, audit and certification. Further, performance measurement also plays an important role in benchmarking. The relation between performance measurement and benchmarking is more complex than the relation between performance measurement and subsidisation, audit and certification.

There is only one way traffic from performance measurement to subsidy and certification in that the latter depend on the outcome of the measurement. Between performance measurement and benchmarking, however, there is two way traffic. On the one hand performance measurement is the foundation of benchmarking, since care providers are compared by means of the outcomes of performance measurement. Benchmarking, on the other hand, is also foundational to performance measurement since requirements, i.e. the contents of the performance norms, are established by comparing the actual results of the care providers. Only when the actual (average) performances of care providers are determined, it is possible to establish a norm. Obviously, the thought behind benchmarking is that these actual (average) performances are the ones one can reasonably demand from all (comparable) care providers.

## 1.2 How do we get from duty of care norms to performance norms?

We have just seen that the legislator instructs the addressees of duty of care norms to both make and apply their own performance norms. The question that will be addressed in this section is exactly how the duty of care norms can be fleshed out in performance norms. Again we will give an example. We will investigate how article 2 of the Quality Act can be made concrete with respect to the field of homecare, more specifically for maternity care.[3]
As a first step a provider of maternity care can formulate the following goal norms in her mission statement. These norms offer a first specification of the general goal of 'sound care':

'The maternity care is delivered timely.'
'The maternity care is given by a professional.'
'The maternity care offered results in a high level of client satisfaction.'

Although these goal norms offer some specification of article 2 of the Quality Act, they are still not concrete enough to enable the care provider or others, such as inspection and patients,

3 The example is taken from Antoinette van Dam, *Kwaliteitsindicatoren in de zorg* ('Quality indicators in the care sector') 2004, <http://www.zbc.nu/main.asp?ChapterID=2616>, site last visited on January 9, 2008.

to determine whether the care provider has in fact fulfilled his duty of care, viz. to offer 'sound care'. Thus, these norms too have to be specified. This can be done, for example, by means of the following norms:

'In *95%* of all maternity care, the professional is present *within 1 hour after the call* of the client.'
'*50%* of all professionals has followed *course X*.'
'*90%* of all clients rates the treatment of the professional with an *8* or higher.'

The latter three rules are examples of performance norms that refer to concrete *verifiable facts* and make use of *numbers*. Only by means of such concrete facts it is truly possible to *measure* the quality of care.
Following Donabedian, most authors distinguish three types of performance norms, viz. norms that contain a *process-indicator,* norms that contain a *structure-indicator*, and norms that contain a *performance- or outcome-indicator.*[4] Although it might seem attractive to focus on outcome-indicators (since these measure the quality of care most directly), it has been argued that performance measurements that make use of all three types of indicators, offer the best quality measurement.[5] In other words, it is better not only to look at the end result, but also at the (alleged or real!) means by which that result has been achieved.
The three examples of performance norms just mentioned contain different types of indicator. The first norm contains a process-indicator: it refers to the process of care (attending the patient within one hour after the call). The second norm contains a structure-indicator that reveals a fact about the way care is organised (courses being followed by professionals). The third norm, finally, contains an indicator that tells us something about the outcome of the process (client satisfaction).

The example of performance norms in the field of maternity care is not unique. In all care sectors, as in many other public and semi-public sectors, performance indicators are being developed and applied in order to give 'hand and feet' to duty of care norms that the legislator has promulgated.

## 2 From performance measurement in business to performance measurement (by order) of the government

### 2.1 Why should lawyers and legal scholars think about performance measurement?

We have seen that the government not only promulgates duty of care norms, but also tells the norm addressees to make and apply their own performance norms. This is an instance of conditioned self-regulation. The term 'self-regulation' refers to the fact that norms are not promulgated by the government, but by citizens (that is to say, not individual citizens but larger societal organisations). The claim that the self-regulation is conditioned points to the fact that these organisations do not promulgate norms from their own free will, but only because the

4 A. Donabedian, The quality of care. How can it be assessed?, JAMA 260 (1988) 12, pp. 1743-1748.
5 R.W.M. Giard, Prestatie-indicatoren als maat voor de kwaliteit van medische zorg: retoriek en realiteit (Performance-indicators as standard for the quality of medical care: rhetoric and reality') , *Ned Tijdschr Geneeskd.,* 149 (2005) 49, pp. 2715-2718.

government obliges them to do so. In fact, in the case of the Quality Act quite some pressure and coercion were necessary before care providers took seriously their obligations to make and use quality systems and performance norms.[6]

The example of the Quality Act shows how the use of performance norms hangs together with, or rather, logically follows from the use of duty of care norms. The example does not, however, answer the question *why* the government has, quite recently, begun to make use of duty of care norms in combination with performance norms instead of classical legal rules that forbid and prescribe behaviour.

Performance measurement is a management technique that has been developed, at the beginning of the 20th century, in the world of *business and industry.* Only since the 1980-ies it has also been applied *to* and *by* the government. Since the end of the 1980-ies performance measurement is applied *to* (semi-) public organisations such as municipality, police and the judicial system. This paper, however, focuses on the fact that performance measurement is nowadays applied *by order of* the government via conditioned self-regulation. Performance measurement is especially demanded from organisations that serve a public interest, such as providers of care and education.

The government, in other words, is doubly involved in performance measurement (governmental organisations are measured and government obliges others to measure themselves). Because of this, performance measurement becomes relevant to legal scholars and scholars of public administration. More specifically, these scholars will have to investigate into the question whether the use of performance norms is legitimate as well as effective. Obviously, when there are doubts about the legitimacy and/or efficacy, it is questionable whether the government should go on with obliging non-governmental organisations to develop and apply these norms. In sections 3 and 4 we will see that there are some serious doubts, in particular about the efficacy of performance measurement. If these doubts hold, it might imply that duty of care norms, or at least the instruction to specify them by means of performance norms, does not fulfil the condition of good legislation.

## 2.2 Performance measurement in business

Before we discuss the efficacy and legitimacy of performance measurement, however, we will investigate how performance measurement entered the public sector. As was stated above, performance measurement finds its origin in business and industry. The period between 1870 and 1910 has been called the Second Industrial Revolution. It is characterised by an ongoing mechanisation of the production, more specifically by the introduction of assembly lines. Another characteristic of this period is the fact that firms begin to invest in (scientific) research to establish more efficient methods of production. Henry Ford (1863-1947) and especially Frederick Winslow Taylor (1856-1915) are generally considered to be 'the' pioneers of the rationalisation of production processes. Taylor's aim was to establish objective production norms and improve, via these norms, the production processes. Henry Ford was the first to bring these insights into practice.

6 E.A. Casparie et al., *Evaluatie Kwaliteitswet Zorginstellingen.* (Assessment of the Dutch Act on the Quality of Health Care Institutions), Den Haag, ZonMw, 2001.

At this time, there was not only approval, but also severe critique of the proposals to automatize and rationalise production processes. As a consequence even today many speak mainly in negative terms about Taylor and 'taylorism'. According to these critical view, taylorism is a worldview that contributes to the alienation of human being from his work. Taylorism, or so the critique goes, reduces man to a cog in the machine. The perhaps most vivid and touching representation of this critique is probably *Modern Times* (1936) a movie in which Charlie Chaplin goes (literally and metaphorically) of the wheels as an assembly line worker.

Despite its rhetorical power, the question remains whether the critique is completely fair. For one thing, Ford en Taylor did not only have the interests of the employer in mind, but also – at least on paper – those of the employees. For example, the first sentence of the first chapter of Taylor's *The Principles of Scientific Management* states: "The principal object of management should be to secure the maximum prosperity for the employer, coupled with the maximum prosperity for each employé."[7]

Whatever one might think of the critique of taylorism, it should not come as an surprise that similar critique is offered from the 1990-ies onward when performance measurement was first introduced in the public sector. According to contemporary critics, the rationalisation of public services undermines the autonomy of the professional and perverts the professional-client relation.

## 2.3 Performance measurement of and by the government

Before we go into the question whether this fundamental critique of performance measurement is correct, we will first have to understand the origin of performance measurement in the public sector. We have just seen that the introduction of performance measurement in the industry can be explained from the ongoing mechanisation of production processes that lend themselves more easily to rationalisation than earlier more traditional production processes. The introduction of performance measurement in the public sector also is a consequence of an ongoing mechanisation, in particular the computerization and the scientific study of public services. It is also connected, however, to the emergence of new ideas about government. These new ideas, called New Public Management (NPM), were introduced in the beginning of the 1980-ies in America and came across the Atlantic at the end of the 1980-ies.

NPM can best be explained by reference to *Reinventing Government* of David Osborne and Ted Gaebler.[8] This book, also called the Bible of NPM, consists of ten propositions that are explicated each in one chapter. The central thesis of the book holds that the government has, especially after WWII, interfered too much in the *execution of public tasks.* This, or so it is argued, has resulted in a cumbersome inefficient and bureaucratic governmental apparatus which blocks innovation and private initiatives. NPM aims to offer an alternative view of government. More particularly, it answers the question how government should operate if not by interfering in the execution of public tasks..
Below, I discuss the ten propositions of Osborne and Gaebler and illustrate them by examples taken from the field of health care in the Netherlands.

7 Frederick Winslow Taylor, *The Principles of Scientific Management*, New York, Harper, 1911.

8 D. Osborne & T. Gaebler, *Reinventing Government*, Adderson-Wesley Publishing Company, Reading MA, 1992.

According to Osborne en Gaebler the government should:

*1. Not row but steer*
The government should formulate policies in general terms, but leave the details to the specific sectors such as care and education. In the first section of this paper, we have seen that this is exactly what the Dutch government has done in health care. Article 2 of the Quality Act only formulates a general duty of care (i.e., the government steers) and instructs in articles 3,4, and 5 care providers to both fill in and fulfil this duty of care (i.e. obliges the organisations to row).

*2. Not serve but empower*
By offering the care providers (and other organisations such as inspection and client organisations) the opportunity to deal with their own business, the government not only gets rid of a lot of work that caused so much bureaucracy, but it also empowers these organisations by offering both the opportunity as well as the responsibility to deal with these tasks in ways they think most efficient and efficacious. This too was illustrated by the example of the Quality Act. The performance norms are not one-sidedly enforced by the government, but are the result of a decision-making process that takes place between inspection and care providers. In other words, the advantage of duty of care norms is that they offer service providers the opportunity and the responsibility to arrange their 'own' business.

*3. Introduce competition*
According to NPM, care providers and other organisations should not only get more room to deal with their own 'business', they should also make use of the 'elbow room' that the government thereby offers them. Among others, care providers should compete with each other, because, or so it is thought, it will lower costs and, at the same time, improve the quality of the services. In other words: competition creates a win-win situation.
This idea too has been realised in Dutch health care sector. Both insurance companies and care providers have to compete with each other. Also discussion has recently started over the question whether the Netherlands should introduce the American P4P ('pay for performance') system. The P4P system is based on the thought that insurance companies giving extra money to the best hospitals, will be an incentive to deliver even better care.[9]

*4. Formulate a mission and goals, not rules*
The next claim of NPM is that the tasks of the service providers are extremely complex and that an optimal execution of their tasks is too unpredictable and too changeable to be put into detailed (legal) rules. Therefore, governments should only formulate a mission and general goal norms, and not prescribe precise behavioural norms.
Again, the Quality Law is an illustration of this credo. Its central article is a duty of care norm and it does not consist of specific behavioural norms.

9 D.J Gouma en O.R.C. Busch, Goede zorg, een kwestie van ervaring ('Good care, a matter of experience'), *Ned Tijdschr Geneesk.*, 151 (2007), pp. 2082-2086.

*5. Choose a result-oriented approach: no input- but output-financing*
So NPM claims that governments should, just like companies, not focus on rules, but rather on results. However, governments should be aware of the risk that mission statements and goals without additional measures, run the risk of remaining empty slogans. One way to prevent this from happening is to assess organisations less in terms of their plans (input), and much more in terms of their results (output), for example via the P4P system.
Hence, we see that the emphasis on result-oriented working and assessing cf. article 3 and 4 of the Quality Act makes performance measurement into an inseparable part of NPM.

*6. Introduce a consumer- instead of bureaucracy-oriented approach*
By analogy with business world, the citizen (in the care sector: the patient) should be perceived as a consumer and customer. Since customer is always right, care providers should no longer operate in a supply-, but in a demand-oriented and demand-driven manner and compete for the favours of the consumer.

*7. Not be spending money, but earning it*
Until the 1970-ies, government did not pay much attention to the budget, but from the 1980-ies onward the budget of the government has gained more attention. NPM argues that, just like companies, government should have a balanced budget. However, this is especially difficult for the health care sector to achieve.

*8. Anticipate: not be curing, but preventing*
A good businessman is not overtaken by the events, but anticipates on new developments. In the same manner, the government should anticipate on, for example, the rise in the ageing population and technological developments and the explosive rise of costs in the health care sector that can be expected to go with these developments.

9. Decentralise and delegate
The central government should delegate more to lower governmental organisations. The central role that has recently been attributed to municipalities in the implementation of the Social Support Act ('Wet Maatschappelijke Ondersteuning') is an example of this.

*10. Introduce a free market system*
Osborne and Gaebler's last claim is that a free market system will result in both better quality and more efficiency.
From the 1980-ies on, a free market system has been a serious topic in the debate about the reorganisation of the health care sector in the Netherlands. In a free market system, improvements in health care are not initiated by the government, but rather by care providers, insurance companies and clients. Care providers will improve the efficacy and efficiency of care, since they have to bid for the favour of insurance companies (compare the P4P system mentioned earlier). The insurance companies in their turn bid for the favour of clients. Clients, finally, can choose care providers and insurance companies by making use of top-100 lists resulting from performance measurement.
Again, there is a quite straightforward relation between a free market system and performance measurement. A free market system demands transparency, and therefore hard figures, and thus performance measurement.

In this section we have expounded, in a nutshell, the tenets of NPM and the role this view has played in the Netherlands. What should have become clear by now is that NPM has been, not only in America, but also in the Netherlands, the driving force behind the introduction of duty of care norms and performance norms. From the 1980-ies on, the government has defended the view that governmental interference with public tasks should be pushed back.
For NPM, in particular the slogans 'not rowing but steering' and 'no rules, but missions and goals', to be successful, classical ways of legislation, viz. behavioural norms, should be abandoned or at least be supplemented by goal norms and duty of care norms.
Moreover, the slogan 'do not serve, but empower' implies that care providers themselves should fill in these duty of care norms. Finally, both for a free market system as well as for adequate supervision 'hard' figures are needed. Therefore the goal norms and duty of care norms should be fleshed out by means of performance norms.

## 3 Legitimacy of performance norms

### 3.1 Introduction

In the foregoing we have both seen what performance measurement consists in and what the historical and ideological roots of performance measurement are. In this and the next section we discuss the question under what conditions performance measurement can be a legitimate way of regulating behaviour. We discuss three kinds of critique of performance measurement of public services.
The first, most fundamental, but also most global point of critique states that public services are principally not suited for performance measurement and that performance measurement has a perverting influence on the public sector. The second argument holds that performance norms are not legitimate insofar as they can conflict with the rule of law. The last point of critique holds that the legitimacy of performance norms is at stake insofar as the efficacy of performance norms can be seriously doubted.
In this section we will discuss the first and second point of critique. In section 4 we investigate the efficacy of performance norms and the implications of a possible lack of efficacy for their legitimacy.

### 3.2 Service providers are unlike factories

According to the first type of critique, performance measurement misunderstands and even distorts the value laden character of public services. The critique holds that service providing organisations do not deliver 'products', but rather value laden services and that for this reason quality assessment cannot take the shape of performance measurement. It is not a coincidence that this critique looks much like the critique that was brought to the fore when performance measurement was introduced in the industrial world. The classical critique of rationalisation of production processes held that rationalisation alienates the workman from his labour and that it reduces him to a cog in the system. In its contemporary shape, the critique says that performance measurement changes the character of service providing organisations as well as of professionals and clients. Government and (semi-)public service providers are wrongly considered to be 'normal', i.e. profit seeking, companies; professionals loose their autonomy; clients become consumers; only efficacy and efficiency matter; and the normative debate

about the values of health, autonomy, solidarity etc. disappears and is replaced by a purely technocratic debate.

The problem about these quite polemic critiques[10] is that there are often worded in very general terms and often are of an ideological and conservative nature. They are ideological in so far as they consider the free market, performance measurement and bureaucracy by definition to be contradictory and thus harmful to public services. They are conservative to the extent that they seem to start from a romantic view of the relation between professional and client as one of trust and mutual respect.[11]

A positive exception to these diatribes is De Bruijn's *Managing Performance in the Public Sector*.[12] He offers a nuanced view of the (dis)advantages of performance measurement of (semi-)public organisations. Part of his objection to performance measurement is practical, part of it is more principled. One of the practical, but nevertheless quite fundamental, objections is the claim that the causal relations between input and output of a service provider are unknown or at least contested. The health care sector is supposed to be 'evidence based', but nevertheless a lot (still) is unknown about the relation between effort and result. The question is, in other words, whether it is fair, and even whether it is possible to judge care providers on their results. This objection can partly be met by taking not only outcome-indicators, but also process- and structure-indicators into account.

Another objection to performance measurement relates to the fact that (semi)public organisations do not deliver their services autonomously, but rather in cooperation with other organisations as well as with their clients. The results of a hospital treatment, for example, depends also on the general practitioner, home care providers and also on patients themselves. This problem too can partly be overcome by not only relying on outcome-, but also on process- and structure-indicators.

As a result of this, it has been observed that performance measurement increases the chance that care providers will try to pass on 'hard cases' to other care providers, or even that they will bluntly refuse them. In the next section we will discuss the risk that care providers will even try to manipulate the figures in order to comply with the performance norms.

We have just argued that care providers do not deal with products but with value-bound services. What does this imply? It is argued that the requirements with respect to products like cookies and lemonade are fairly univocal: they have to be safe, not too unhealthy, tasty, etc. What, however, are the goals of service providing organisations like care providers? Guar-

10 See for example Eliot Freidson, *Professionalism. The Third Logic*, Cambridge, Polity Press, 2001. Evelien Tonkens*, Mondige burgers getemde professionals. Marktwerking, vraagsturing en professionaliteit in de publieke sector,* (Mature citizens, tamed professionals. Free market, demand-steering and professionalism in the public sector) 2003 Utrecht, NIZW. Gabriel van den Brink, Thijs Jansen en Dorien Pessers (red). *Beroepszeer. Waarom Nederland niet goed werkt* (Professional pain or honour. Why the Netherlands does not work well) Boom, Amsterdam, 2005.

11 Note, however, that Tonkens is also in favour of inter- and supervision and more qualitative forms of inspection.

12 Hans de Bruijn, *Managing performance in the public sector*, London, Routledge, 2002. (Translation of *Prestatiemeting in de publieke sector. Tussen professie en verantwoording.* Utrecht, Lemma, 2001.) Also see Hans de Bruijn, Management van professionals: tussen beroepszeer en beroepseer (Management of professionals: between professional pain and professional honour), *Trema,* November 2006, pp. 433-438 and Hans de Bruijn and Jan van Helden, Effectief prestatiemanagement bij professionele organisaties in de publieke sector (Effective performance management in professional organisations in the public sector), *Overheidsmanagement*, 2007, nr 1, pp. 20-26.

ding and advancing health care and wellbeing is one important goal, but not the only one. The autonomy of the patient is a central value too and fair distribution of care (implying that 'hard cases' should not be refused) is again another aim. Also, care providers have to comply with requirements that have to do with the environment and working conditions for example. Now, the question is: how can we measure, given these very different and possibly conflicting demands, whether 'good care' has been offered? What if the patient is cured, but his autonomy was not fully respected? What if the quality of care was good, but the distribution was unfair? What if working conditions were not met, etc.? The assessment, in other words, is not so much a matter of facts, but of values.

The question here is not only *how* the assessment should be made, the question is, first and foremost, *who* should determine what aspects should be measured and *who* should determine how these different aspects should be weighed. The most fundamental question is whether such a fundamental assessment can be made via performance norms that are not made by a democratically chosen legislator, but by organisations from the health care sector.
There is yet another problem connected to the fact that performance norms are not promulgated by the legislator. The main justification for the legislator to use goal norms, duty of care norms and other so-called open norms is that society is complex and changing constantly. Therefore a legal system that only consists of clearly specified behavioural norms will always lag behind the facts. However, the performance norms that the care providers are supposed to make and apply, suffer from the same deficit as traditional behavioural norms. These norms too need a fairly stable environment in order to function well. In other words, by shifting the obligation to make more precise (performance) norms from the government to societal organisations, the legislator also puts the burden on them to constantly update these norms in order to keep step with changes that take place.

### 3.3 Performance norms and legal principles

The objection to the fact that performance norms are created and enforced via conditioned self regulation brings us to a second objection to performance norms. The question to be discussed is: does this type of conditioned self regulation conflict with the rule of law?
It should first be noted that performance norms do not satisfy the demand of legality. One can respond to this objection that this is not a serious objection in the Dutch democratic system which is characterised by its emphasis on consultation and consensus for thereby at least the demands of inclusiveness and representation will nevertheless be met. In health care, for example, performance norms are not determined by the care providers themselves, but by a coordinating organisation which consists of delegates of the Inspection and several kinds of organisations from the health care sector. Of course, one can have doubts about the extent to which these delegates truly know about the work being done on the 'shop floor', i.e. to what extent they truly represent professionals and patients. However, the demands of inclusiveness and representation are met at least to some extent.

A more important question is to what extent the demands of legal certainty and equality are fulfilled. At first sight, one might think that performance norms truly serve the aims of certainty and equality. They serve certainty since they make more precise what the rather vague duty of care norm of the Quality Act consists in. However, performance norms turn out to be

less univocal and stable than one might think since there is a tendency to constantly re-adjust performance norms.

Performance norms seem to fulfil the demand of equality. On the one hand they formulate clear and univocal demands that make possible not only to treat like care providers alike, but also to treat unlike care providers unalike. For example, performance norms for hospitals differ from norms for nursing homes or rehabilitation centres. As long as the differentiations are approved of by all stakeholders, it seems that the principle of equality is served by the introduction of performance norms.

However, and now we come tot the most critical point, for performance norms to truly fulfil the demands of equality and certainty, it is crucial that performance norms are sound. This implies, firstly, that performance norms should measure what we want them to measure and not something else. Secondly and at least as important is the demand that it is possible to apply the norms properly and that is impossible or at least not too easy, to apply them wrongly. This, in turn, implies that addressees should not have (too many) opportunities to manipulate the outcomes of the measurement.

In the next section we will learn that there some serious questions can be posed with respect to both demands. It turns out to be difficult to make good performance norms and it turns out to be possible and (obviously) attractive to manipulate the figures.

## 4 Efficacy and legitimacy of performance norms

### 4.1 Goals and audiences of performance measurement

In section 1 we have seen that performance measurement can be used for different *purposes*. The primary goal is to make transparent how service providers have performed. The services that have been delivered can then be compared to the services as intended. If there is a discrepancy between the delivered and the intended services, two things can be done. The purpose can be that addressees learn from their failure and try to improve their performances. Another purpose can be that addressees give account of their performances and that they, if necessary, get fines or rewards for their insufficient or excellent results. The thought behind the second aim is, obviously, that punishment and reward are an indirect but effective route to improvement. In other words, transparency, learning and improving, accountability and fines and rewards, are the keywords of performance measurement.
A second distinction is the one between internal and external *audiences*. It should be noted that the distinction is relative. The board of a hospital for instance, is external in relation to a specific department of the hospital, but it is internal in relation to the inspection. The distinction is used to distinguish between groups of persons with and without detailed information about the organisation. Members of the internal forum know, so to speak, the world behind the numbers and figures, whereas members of the external forum don't.

Members of the external forum only 'see' the hard figures (what percentage of maternal caretakers arrived within one hour after the call, how many followed a course, etc.). The

external audience cannot, at least not without further explanation and clarification by the care providers, interpret and, if necessary, nuance the outcomes. A good example of such an external audience is the class of potential patients/consumers who can acquire information about hospitals via different top-100's that yearly are published. A significant detail about these top 100's is that in 2006 there was no concurrence whatsoever between the three most prominent rankings.[13]

A serious risk of performance measurement is that it is often introduced to serve an internal audience with the primary aim of quality improvement, but that at some point external audiences also get the disposal of this information which is then used for the purpose of accountability, benchmarking, and even for 'naming en shaming'. Shifts of purposes and audiences is risky for several reasons. In the first place, external audiences cannot properly read, i.e. interpret and nuance the outcomes. External audiences are more inclined to treat the outcomes as 'hard figures' than as the indicators they were intended to be. Since services providers are aware of the risk that internal figures might be used externally, they will not be fully open and they will be inclined to give a rosy picture of how things are in the organisation. Investigations on this point show that both risks are real.[14]

We have seen that NPM favours competition between service providers, but it has been show that the more competitive (public) service providers have to be, the less willing they are to share their 'formula for success'. Schools and hospitals for example, are inclined to show *that* they are successful, but not *how* they achieve the success. This is especially so when they are rewarded (P4P) for excellent results. Another effect of competition based performance measurement is that public service providers are less inclined to be inventive. When the evaluation is good, it is safer to keep what you have than to try new but risky routes. Also, organisations sometimes only improve those parts that are measured and do so at a cost of parts that are not measured.[15]

The conclusion of this section is that if performance measurement is used externally for the purpose of benchmarking and competition and sometimes even for naming and shaming, the risks of improper use and of withholding and even manipulation of facts increases.

## 4.2 Is it possible to make good performance norms?

Before we discuss an example of manipulation of performance measurement, we first have to find out whether it is possible in principle to make sound performance norms. For performance norms to be sound, they should be valid, reliable, responsive and useful. This section briefly discusses these requirements. We will use the example of hospital mortality to illustrate them.[16]

13 R.W.M. Giard, Ziekenhuizentop-100: wisselende ranglijsten, wisselende reputaties ('Hospital top-100: changing rankings, changing reputations'), *Ned Tijdschr Geneeskd.,* 150 (2006) 43, pp. 2355-2358.

14 See e.g. Hans de Bruijn, 2001 en Gwyn Bevan & Christopher Hood, What's measured is what matters: targets and gaming in the English Public Health System, *Public Administration* vol 84 no 3 2006, pp. 517-538.

15 See De Bruijn 2001.

16 I take this example from P.P.M. Harteloh en A.F.Casparie, *Kwaliteit van zorg. Van een zorginhoudelijke naar een bedrijfskundige aanpak* (Quality of care. From an approach to the content of care to the management of care), Maarssen, Elsevier/De Tijdstroom, 1998 (4e druk).

*1 Validity*
The first requirement is that indicators are valid. Valid implies that indicators measure what we want them to measure. So in our example the question is: is hospital mortality a valid indicator for the quality of care? In order to answer that question we have to make a distinction between sensitivity and specificity.

*1a Sensitivity*
Harteloh en Casparie argue that only 11% of the cases of low quality of care can be signalled by means of the indicator hospital mortality. The indicator has a low sensitivity because it does not allow us to detect low quality that is caused by other facts, such as e.g. infection en re-operation. In other words: low sensitivity implies in this example that the indicator does *not* detect 89% of the cases of low quality. This implies that we need other indicators if we want to properly measure the quality of care.

*1b Specificity*
Hospital mortality turns out to be a highly specific indicator. In 96% of the hospitals where the quality is good, the mortality is low. High specifity implies, in other words, that only few hospitals are wrongly accused of offering low quality. One might say that mortality is a 'fair' indicator. As the example shows, the final judgement about the validity of an indicator is not a hard and objective fact. It is rather the outcome of a normative weighing and balancing of specificity and sensitivity. Depending on the decision that it is more important to find 'bad' hospitals or rather not to falsely accuse 'good' hospitals of delivering bad care, the verdict about the validity of mortality will be negative or positive.

*2 Reliability*
An indicator should not only be valid, but also reliable. Reliability means that all institutions should register their facts in the same manner. Again when we look at our example, we see that it is very reliable. There is hardly ever debate about whether someone is dead or not, it is not hard to register 100% of the cases and (not completely unimportant) it is hard to manipulate the mortality figures since all deaths have to be reported anyhow.[17]

*3 Responsiveness*
Not only should the indicator be valid and reliable, it should also be responsive. This implies that an indicator should also serve the purpose of reflecting or even initiating quality improvement. Hospital mortality has turned out not to be suitable for this purpose. In the USA no changes in mortality figures have been found despite drastic changes in health care.

*4 Usefulness*
The usefulness of an indicator depends among others on the accessibility of the information needed, the costs to collect and process the information and the acceptance of the indicator by all stakeholders.

17 However, even here there is an escape: according to the definition of hospital mortality all patients who die within 30 days after having left the hospital should also be taken into account.

*4a Accessibility of information*
The accessibility is not an issue in the case of mortality figures. In other cases, however, this can be a more serious issue.

*4b Costs*
The costs of collecting and processing are not a trivial issue. Many organisations complain about administrative work load and the number of hours not spend on their core task, offering cure and care. So many openly doubt whether the costs of performance measurement are outweighed by the benefits.

*4c Acceptance*
The acceptance of the indicator by addressees is important. In the USA for example, there has been a discussion about the soundness of the indicator hospital mortality. Partly because of this discussion, publication of hospital mortality numbers has been stopped in the USA in 1995.

The purpose of this brief discussion of the requirements for good indicators was merely to show that one (i.e., the legislator) should not think to light heartedly about the construction of good indicators. It is often referred to as 'merely' a practical problem, but it turns out to be problem that presupposes normative appraisals and takes a lot of time and money.

## 4.3 Perversion

The example of mortality in hospitals shows that the construction of indicators that satisfy all methodological criteria is not a simple matter. In this section we will discuss the fact that even if it is possible to make good indicators, there still is a discrepancy between what they actually measure and what they intend to measure. This discrepancy need not be problematic as long as indicators are taken for what they are, viz. literally 'indicators' of the quality of the service being offered. Now as we have seen, one of the risks of the use of performance norms is that they are not taken to be indicators of, but as hard evidence for, the quality of care.

If an indicator only has a signal function, then a (good or bad) outcome can be a reason to investigate possible causes of the outcome. In such cases the outcome will be discussed in a dialogue between investigator and the investigated organisation. For example: does high mortality indicate bad quality, or is it to be explained from a flue epidemic, a heat wave, or from the fact that the hospital's clients are atypical as compared to other hospitals?

However, as soon as indicators are treated as a hard figures ('dials') that can not be debated and negotiated, the figures will start to lead a life of their own and it will be hard to correct wrong interpretations. Obviously, this is a serious problem both from the perspective of supplying objective information to the audiences, as well as from the point of view of justice to the organisations under investigation. The problem, however, is that there is an ineradicable tendency to treat numbers as hard facts, even if they were originally intended to serve a signalling function only.

This brings us to the problem mentioned earlier: the risk that organisations will manipulate the outcomes of performance measurement.

Bevan en Hood investigated several health care institutions in the English National Health System.[18] According to Bevan and Hood the English system is more directed at punishment and reward and 'naming and shaming' than other British countries (and also than the Netherlands it seems). Bevan and Hood claim that, as a consequence, English institutions deliberately distort and embellish their performances.

One of the (almost hilarious) examples they discuss deals with the performance norms for the waiting period at the Accident en Emergency Department. One of the norms is that a patient should be treated within 4 hours after arrival at the department. According to Bevan and Hood this specific performance norm causes four different kinds of 'output-distorting gaming'. First, a study of waiting time distribution revealed a frequency peak around 4 hours, whereas earlier there was no such peak. Second, departments drafted in extra personnel and third they cancelled operations in the period over which performance was measured. Fourthly, patients had to wait in queues of ambulances outside the A&E Department until the department was confident that the patient could be seen within four hours. Bevan and Hood claim that this last practice may have caused (possibly fatal) delays to other patients when ambulances (that were in fact available) were waiting outside the hospital to offload their patients.

Since Bevan and Hood studied the effects of performance measurement in English health care facilities where accountability and even 'naming and shaming' play a vital role, we cannot draw any straightforward conclusions when it comes to evaluating the efficacy and legitimacy of performance measurement of health care in the Netherlands. However, even if only a fraction of the consequences they discuss also occur in the Netherlands, that would have implications both for the efficacy and the legitimacy of performance measurement.

## 5 Conclusion

Must the conclusion be that performance measurement is a very dangerous instrument that we should abandon and abolish as soon as possible? Although Bevan and Hood are critical about performance measurement, they nevertheless do not draw this radical conclusion. Their prime concern is to disconnect performance measurement from naming and shaming.
Another warning is to be careful about changes in goals and audiences of performance measurement. Changes (from internal to external audience and from learning to accountability) are possible, but hey should be discussed openly with all stakeholders. Moreover, if necessary indicators should be adapted to the new goals and audiences. Also, one should be careful not to focus one-sidedly on outcome-indicators at the cost of structure- and process-indicators. A final warning is not to introduce performance measurement too light heartedly, for it is more easy to introduce than to abandon it.[19]

18 Gwyn Bevan & Christopher Hood, What's measured is what matters: targets and gaming in the English Public Health System, *Public Administration* vol 84 no 3 2006, pp. 517-538.

19 De Bruijn, 2001.

The main conclusion is that in the next few years, the effects of performance measurement should be studied very carefully. Does it result, as critics claim, in a more unevenly distributed and scantier care? Or does it result, as proponents hold, in an increase of both quality and efficiency? To answer *this* question, we need something other than the ideological pleas of critics and adherents. Instead, empirical research is needed. A meta-analysis of the Dutch Health Care Council shows that there is hardly any research into the effects of performance measurement.[20] Therefore we should not only measure the quality of (semi-)public services, but we should also start to empirically investigate the quality and the side-effects of performance measurement.

As long as empirical evidence is lacking, legislators should not think too light heartedly of regulation via a combination of legal duty of care norms and self regulated performance norms.

20 Gezondheidsraad/Raad voor de Volksgezondheid en Zorg, *Vertrouwen in verantwoorde zorg? Effecten van en morele vragen bij het gebruik van prestatie-indicatoren.* (Trust in sound care? Effects of and moral questions about the use of performance indicators), Signalering ethiek en gezondheid 2006/1, Den Haag, Centrum voor ethiek en gezondheid, 2006.