

01 1994 This bulletin aims at informing the international forum on recent judicial developments. It is issued by the Dutch Ministry of Justice, Research and Documentation Centre (director: dr. J. Junger-Tas), P.O. Box 20301, 2500 EH The Hague. Tel.: 31-703707610; nr. P.B.A. ter Veer (ed.), Fax: 31-703707948. To subscribe to this bulletin (free of charge) write to this address. Design: H. Meiboom; printed by SDU, The Hague; production: M.E. Samplimon (RDC).

09 Dutch penal law and policy Notes on criminological research from the Research and Documentation

Centre Ministry of Justice The Hague The Netherlands

Measuring police performances by G.J. Terlouw, M. Kruissink and C.J. Wiebrens

At the end of 1993, a large-scale reorganization of the Dutch police will be completed. The existing 148 municipal police forces and 17 state police districts will then be integrated into 25 regional forces and one national force. The function of the latter will be to support the regional forces. Together with the reorganization a new financing system for the police is being introduced. The old system of financing input, simply meaning that a police force is supplied with a fixed sum

of money, may eventually be replaced by a system of partial output financing. Output financing implies that part of the funds to be supplied to the force is dependent on the performance of that force. A requirement of output financing is that insight into the performance of police forces is needed in the sense of (absolute or relative) ratings. Ideally, these ratings should also give indications about the factors that are responsible for the performance level achieved.

Next to this change of financing system as an argument for developing a measurement instrument for police performance, police authorities like the public prosecution office and (regional) government request for more tangible information on police performance.

'Hands-off management'

In developing the design proposed below for measuring police performances we departed from two starting points. The first is a managerial model, in which roles and responsibilities for the different players are defined. It uses the idea of 'steering at a distance' or 'hands-off management'. The basic thought here, is that there are limitations to the exchange of information on optimizing production processes. In other words, it is not very efficient to have local police forces basically managed largely from a ministry. This leads to the conclusion that those forces should be relatively autonomous. They are responsible for achieving the necessary results and how they do that is their own business. The government in its turn, is responsible for things like furnishing funds required for realizing the plans and exercising (quality) control on the execution of those plans:

This idea is reflected in the so-called 'principal-agent' relationship. What the police should do, to what effect, what the police priorities and posteriorities are, is decided by the principals, who set the police agenda. The main principals are the police departments (the Ministry of Justice and the Ministry of the Interior), regional and local administration, and public prosecution.

The chiefs of the (after the reorganization) regional forces can be considered the agent in the relationship. They are responsible for the output the police attain and for the management of the police force.

The principals decide what results they would like to see, and have to reach an agreement with the police force management on what has to be done (e.g. in an annual projection). The agent has, through an output/effect report, to render account of what has been achieved in relation to the agreements reached. This report should inform the principals on the efficiency and appropriateness with which the means provided were used. The implication is that an adequate registration of performances (qualitative as well as quantitative) is important and a necessity.

About input and output

The other point of departure is the well-known input-throughput-output-effects chain.

Input is formed by the means provided by the Ministry to the police force for the period of one year. We can express input in units like dollars, personnel or time etcetera. The unit per se is not too important, as long as most or all parts in the model (input, throughput, output etcetera) use the same units. We opted for time as a common unit of measurement. That yields an input

consisting of the total number of man-hours per year available to a police force.

Throughput is what happens in the police organization itself. Throughput stands for the process of converting the furnished means into products, or the production process. The production process is controlled and steered by the police management.

Output is the production of a police force. If we want to be able to make statements about the performance of the police force, it is required that we somehow link output to input. We create this link in several steps, starting by dividing output into products and services (explained below).

Effects are the desired consequences of output realized on the clients intended, for instance a general feeling in the public that it is safe to walk the streets at night.

Products and services

As stated, we divide output into products and services. Products are countable and relatively easily measurable, consisting of crimes cleared through arrests, fines, licenses etcetera. We also use the term 'hard output' here. The possible effects of hard output, e.g. less burglaries because notorious burglars are imprisoned, are not (yet) taken into account in the model. These effects could be extremely diverse e.g. less burglaries might in the long run also lead to less fear of crime under the population and/or to diminishing crime levels in other fields than just burglary.

Activities such as giving advice, patrolling, prevention etcetera are considered police services. 'Soft output' like services is considerably harder to measure or count than hard output, because the matter is less concrete. A possible way to get an impression of the quality of services though, is to try and measure their effects or impact. Examples are for instance the population's judgement about the police, or the perceived chance of victimization. Checking consequences of police services in order to get a picture of the quality of the services realized by a police force thus means that the information on soft output is obtained by indirect measurement.

Linking output to input

The next step is the linkage of the hard output indicators to input. The hard output indicators (recorded crimes) are to be expressed in the same units we used for input, that is in hours. To achieve that, we apply the so-called 'standard times' dating from the 1988 reallocation research project.¹ A standard time is defined as the average time necessary to handle an incident (burglary, shoplifting etcetera) completely. By multiplying the incidents in each crime category in the police administrations with their respective standard time, it is possible to calculate the number of hours spent on hard output per force. This also allows us to express the number of hours spent on output as a

¹ For a description, see Wiebrens, 1990.

Figure 1: The performance matrix

clearance time	high	A1	A2	A3
	middle	B1	B2	B3
	low	C1	C2	C3
		A		
		D, F, G, H, I		B, C, E
		N	L, M	J, K
		high	middle	low
		workload		

percentage of input (total number of hours available per force), since the input is known of course. The output-input percentage calculated in this way would not be a perfect indicator of police performance or productivity, however. Police forces necessarily have to spend time on registering crimes and incidents reported to them. Relatively speaking, some police forces have to deal with many more crimes and incidents than other forces, which means some forces 'lose' much more time in administrative processes than others. In other words, forces differ in workload: the number of crimes reported per officer. These differences in workload between police forces have to be taken into account if we want to get an accurate estimate of performance. It is our assumption that forces that cope with a high workload are limited in their development of extra activities next to recording crimes and that forces experiencing a relatively low workload should be in a position to generate a lot of actions aimed at clearing up crimes, traffic control etcetera. Below we will give a real-world example of the measurement of police performances, at least where these relate to hard output. Then, we will explain how the information on the returns of police services, or soft output, fits into the product performance model.

Measuring police performance in practice

Police products

Based on their 1989 administrations, we measured the hard output of fourteen police forces. The data concern the number of recorded and cleared up crimes, fines etcetera per force for that year. In order to construct a performance indicator, we set out on calculating 'intime' and 'outtime'. *Intime* is the average time necessary to register a crime or incident without delivering

a final product like an arrest or fine. *Outtime* is the average time necessary to deliver a final product with regard to an incident or crime after the crime or incident has been recorded.² The cases recorded for each crime category in the police administration are multiplied with their appropriate intime and outtime (provided the case has been solved). Subsequently the intimes and outtimes calculated for each crime category are summed, resulting in a total intime and outtime figure. Together, total intime and outtime give the time invested in hard output. Next, we related the total intime (the time needed for recording all reported crimes) to total time available per force (input), which gives us an idea of the workload. The fourteen forces were grouped into three workload categories (high, average and low workload).³ The total outtime per force was related to the sum of total intime and outtime per force, resulting in a number that indicates what fraction of the hard output time was spent effectively, that is resulting in end products. This indicator is called *clearance time*. Like workload, three categories for clearance time were created, specifying forces which spend a relatively large, intermediate and small fraction of their hard output time on products.

The results of these calculations are plotted in a so-called performance matrix. The horizontal axis of the matrix represents the workload categories, while the vertical axis holds the outtime categories. Each of the fourteen police forces, indicated by the letters A through N, can be placed into one of the nine resulting cells (see figure 1). Earlier we stated that we assume that forces which have to spend a relatively large amount of their time on the administration of reported crimes, incidents etcetera (forces with a high workload), have relatively little time left for clearing up crimes, arresting suspects etcetera. For that reason we expect these forces to be plotted in the bottom left cell (C1) of the performance matrix. On the other hand, police forces that have a relatively low workload can spend relatively more time on clearing up crimes etcetera. So we might expect these forces to show up in the upper right cell (A3). Forces with an average workload should achieve an average outtime (cell B2).

The diagonal from bottom left to upper right (C1-B2-A3) shows the expected positioning (or performance) of forces at a given workload. Superior forces are those which show up above the diagonal (in one of the upper left cells A1, A2 or B1), performing above expectation. Those which take a position under the diagonal (in cells B3, C2 or C3) have a performance level that is below expectation. Six of the fourteen police forces are placed on the diagonal C1-B2-A3 in figure 1 (N, D, F, G, H and I). Their clearance times are as expected, given their workload, and so they display an average performance level. Force A is performing above expectations, combining a high clearance time with an average workload. In the right corner of the matrix are seven forces (B, C, E, L, M, J and K), whose clearance times are lower than could be expected considering their workload.

It is important to be aware however, that in the example

² Intime for crime category X plus outtime for crime category X is standard time for crime category X.

³ Each category covers one third of the range from lowest to highest score. This way of classifying is used for all other indicators in this paper as well.

Figure 2: Classification of fourteen police forces by the effects of their services

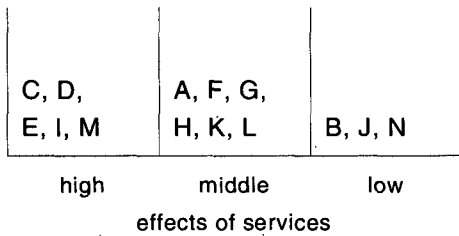
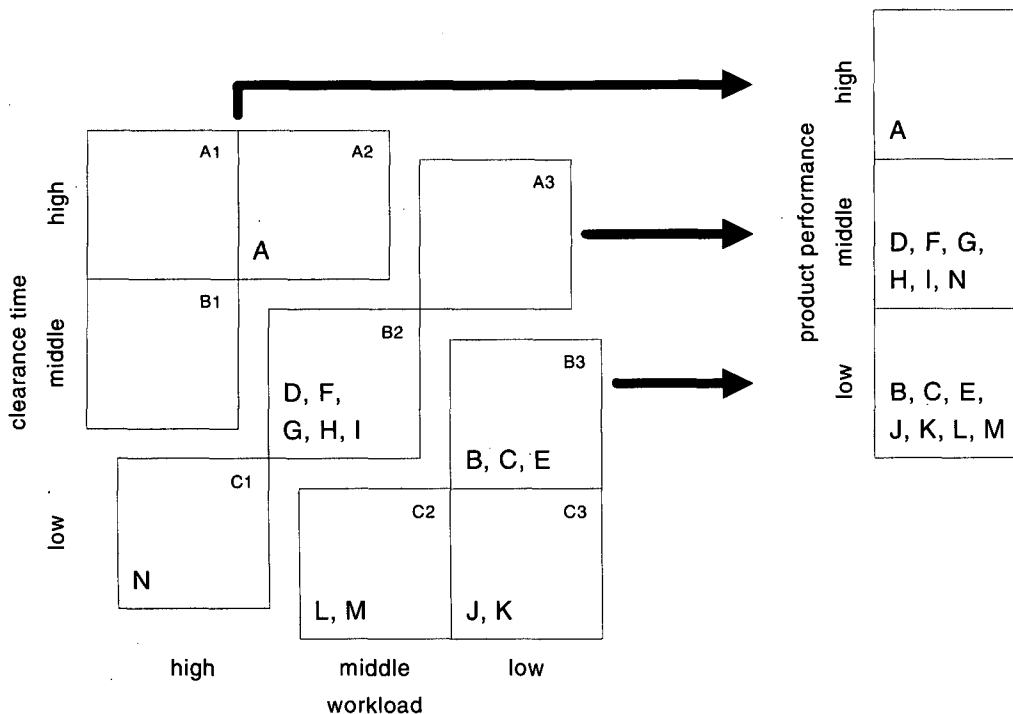


Figure 3: Classification by position in the performance matrix



the expected performance is based on a calculated average. Principals may well prefer other positions in the matrix as a norm, e.g. above the centre diagonal (diagonal B1-A2). Selecting the criterion is a policy issue here.

Police services

The performance matrix depicted in figure 1 is an incomplete representation of police achievements, as it only covers police products (hard output). The other side of police output, services, is not incorporated.

As suggested earlier, a way to get an estimate of the output level on police services is by polling the population on police services. To acquire this information, we used data of fourteen local (standardized) crime surveys, pertaining to the fourteen forces used in the earlier example (Van Dijk, 1991). We computed scores for five aspects: judgement of the public on overall police performance, police availability, willingness to report crimes, victimization rates and safety feelings. These scores were combined into one indicator as a measurement of police services, which has three categories: high, medium and low. For the moment,

each aspect has the same weight in this combined indicator. Figure 2 shows the ordering of the fourteen forces according to their relative scores on the service indicator.

Combining hard and soft indicators

Now, we would of course like to have a look at the combined indicators for police performance on hard output and services. In order to be able to do that we first grouped forces according to their position in the product performance matrix (figure 1) into one of three classes: forces that perform better than expected, those that perform as expected, and forces whose performance is below expectation. This manipulation is shown in figure 3. Essentially, what happens is that the performance matrix is collapsed into a one-dimensional performance scale. The last step consists of merging the classifications for performance on hard and soft output into one new matrix. The horizontal axis represents the performance of the police forces on services, as measured by the effects of services, the vertical axis represents the performance of forces with regard to products (figure 4).

Figure 4: Matrix product performances versus effects of services

		A1	A2	A3
high			A	
		B1	B2	B3
middle	D, I		F, G, H	N
		C1	C2	C3
low	C, E, M		K, L	B, J
	high	middle	low	
	effects of services			

Comparing police forces

The last matrix provides insight into the performance of forces with regard to products and services. Just like we did for the product performance matrix, here too we assume that a position in one of the cells on the diagonal C1-B2-A3 is an average, an expected position. Police forces plotted in one of the cells above this diagonal in general perform better than other forces. They invest relatively much of their time in effectively clearing up crimes etcetera, while at the same time they manage to provide valued services to the public. Forces located under the diagonal perform lower than expected on products as well as services. And again, other criteria can be chosen by principals as a minimum performance level to be achieved.

From figure 4 we note that two police forces (B and J) show a relatively low performance on hard output while the returns of the services they delivered are also less than what we find for the remaining forces. Forces K and L achieve average results on services, their hard output is low though. 'We're not doing so well on solving crimes, but we're very good in the field of surveillance and prevention' is a widely used excuse in police circles. The matrix shows that this defense is valid for 3 forces (C, E and M), that, though they score relatively low on product performance, achieve above average results with regard to services. Police force N is average on product performance, but its service performance is relatively low, meaning that the overall performance is below average. Average ratings on both products and services are obtained by forces F, G, and H.

Three forces display superior performances. Force A is above average on hard output in combination with average results on the success of provided services. For D and I the situation is the opposite: they combine an average product performance with relatively high returns on soft output.

Uses, limitations and prospects

The previous paragraph showed how the performance model can be used to compare the performance of different police forces in a specific year. However, there are several other ways in which the proposed design may be used.

The model presented groups all police forces into one matrix, irrespective of geographical or other external circumstances. While it might be argued that the place of a police force in the performance matrix also reflects to what extent factors like large numbers of commuters, tourists etcetera are accounted for by that force, at the same time the variability in circumstances like these is often large and their direction unpredictable. In special cases, we could therefore consider selecting police forces that work in the same type of circumstances in order to calculate a performance matrix for comparing those forces.

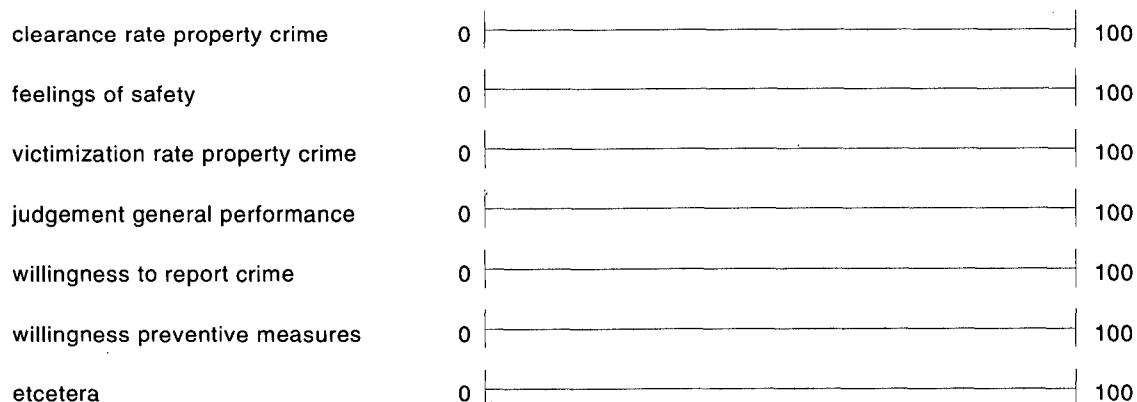
Another way of using the model, is to compare performances within forces over time, that is, look at the development in performance for a specific police force over months or years. If we would do that at the macro level, trends may be analyzed. This approach also offers opportunities to put performance developments in perspective. For instance, a five percent performance improvement in a year is not so impressive anymore if one knows that the average performance increase (over all forces) is four percent in the same year.

Furthermore, we could plot the performances achieved by forces in, say, 1991 into the 1990 matrix. The thought behind this idea is that management decisions would be taken with respect to the 1990 place in the matrix, while after a year, the police force is judged on their place in the 1991 matrix. The definition of the 1991 matrix's cells is different however from that of the 1990 matrix's cells, since the scores of the police forces on workload, clearance time etcetera (which define the range and as a consequence, when classifying, the cell limits) differ from year to year. In 'normal' use, we would judge the performance of a force in 1990 relative to the performance of the other forces in 1990, and in 1991 relative to the 1991 performances of the others. These are all judgements based on relative achievements.

Plotting the 1991 performance in the 1990 matrix would give us an idea of the absolute change in performances, and also whether the changes took place according to or along the lines of the management decisions made based on the police department's place in the 1990 matrix. This brings us to the subject of the specificity level of the model. In the form presented above, the performance matrix shows an overall picture: though weighted according to their standard times, different crimes and services are simply added and are each represented by just one number. This simplicity is an attractive feature of the model, but it has disadvantages as well. If we want to know in what direction changes take place, the strengths and weaknesses in the performance of forces, we really need a more differentiated image.

For this, the current model offers the opportunity of *zooming*: it is possible to calculate separate matrices for the component parts of the general performance matrix, that is matrices for e.g. property crime, violence and so on.

Figure 5: Performance profile, example



Zooming in even closer would result in matrices for the components of (e.g.) property crime, which means we would get matrices for the separate crime types like burglary. Matrices like these give detailed information on the magnitude and directional changes in performance taking place and could be very helpful in managing the police force. For each component matrix a specific norm concerning the desired performance level may be defined. Of course, in turn the component matrices again offer the possibilities of comparison between forces and to own performances in earlier years.

Next to this 'performance' zooming from a general performance measurement of hard and soft output to more detailed component matrices, the concept of zooming can also be applied geographically: instead of comparing municipal police forces (or, in the new police organization, regional forces) with the average national performance level, this can also be done at the regional level, comparing the districts within the region, or maybe even within one police force, comparing different divisions. Both ways of zooming can be used at the same time.

Another way of refining the model – especially with respect to assessing performance on services – would be to add to it what we call 'performance profiles'. A profile is an index or scale which represents the performance (output or effect) on a specific aspect of police work, say victimization rate. Figure 5 shows some examples of what profiles might be like.

Performance profiles like those presented in figure 5 would enable us to compare forces directly at a considerably detailed level. Besides that, problems like determining which indicators should make up soft output, assessing their weights and trying to combine information of considerably diverse character which at the same time could be correlated, into one concept of soft output, are – at least at the measurement level – avoided, because the separate indexes are not combined. For every profile we can of course calculate an average 'expected' score on the scale, to which individual performances can be compared. The problems of integrating and weighting different information are shifted to the political level though, where the decisions concerning which performance aspects weigh heavily and which are less important still have to be made in order to be able to make accountable

management decisions. And, as more details are added to the model, more of its initial attractive simplicity is lost. We may end up with a basic performance measurement model that has different shapes, each tailored to the specific situation in which it is used. The 'top' or global overall model, is composed of all the elements which may show up in the more detailed versions of the model. The Ministry of Justice and the Ministry of the Interior, being the principals managing 'hands-off', might be expected to be satisfied with an overall view of the situation, that is with global matrices like the one presented in figure 4. Chiefs of police on the other hand, the managers of the force, would probably prefer more detailed information: they would want to know if their performance is up to or above par and what aspects of their performance are substandard, in order to be able to make decisions on a change of course.

Anyway, whatever the definitive shape of the model will be like, there are several basic problems to be solved. An important issue is the validity and reliability of the data that constitute the input of the model. Standards have to be designed for measuring, recording and categorizing incidents and crimes reported and solved. Since the standard 'intime' and 'outtime' change in the course of time due to e.g. technical developments, they should be calibrated regularly if we want to have reliable gauges. Another class of problems is formed by what was already briefly touched upon in the previous paragraph, namely the question of what we should and need not measure in order to get a reliable approximation of how the police are performing. The fact that police activities are multifaceted and that the objectives of policing are largely unspecified complicates this issue. Preferably, a particular police goal should determine the performance indicator to be used. This matter is confounded even further by the fact that we can not tell what interactions exist between hard/soft output and their respective impact. To give an example, if the police solve a lot of burglaries, and arrest many burglars, the effect of this hard output may be that the number of burglaries diminishes. But this hard output could also influence the evaluation of police services (soft output) by the public: people feel safer and may give their police force better performance ratings. On the other hand, these changes could also be caused in part by better

surveillance, or by simultaneous campaigning for the installation of better locks on doors to secure apartments against the risk of being broken into. This essentially illustrates the problem of correlated indicators mentioned earlier.

Even when effects are being measured, it is hard to assess the contribution of the police. It is impossible to isolate the impact of policing on the community, as other institutions, environmental and social circumstances (e.g. unemployment) influence society as well. Their influence extends to those aspects of society that we use for the measurement of police performance, like feelings of insecurity or crime rates. One might expect that a region where retired people are overrepresented will score relatively high on feelings of insecurity. Cities with lots of pubs and cinema's etcetera will show higher crime rates than other cities. Police activity is just one of many contributors to an existing situation.

The final shape of the performance model will also be determined by the issue of assessing the weights of the component parts of soft output (like police availability, willingness to report crimes) before merging them into one indicator. Additionally, we face the issue whether those component indicators of soft output can be combined anyway, since the concepts being measured are so different. Should we be satisfied with profile sketches? Next to these considerations, the orthogonality of the X and Y axes of the matrix suggests an independency which, considering the examples given above, is dubious or non-existent. Also, the units of measurement used for each axis differ: time for hard output and combined scale scores for police services.

Creating an acceptable and reasonably reliable police performance measurement instrument promises to be a difficult project, in which the plan explained above forms the initial step. However, on the basis of the design presented here, it should in our opinion be feasible to arrive at a design that yields good estimates of police performance and which thus enables us to compare police forces on several levels. We expect the outcomes of future 'real-life' tests as well as further theoretical development to help us in adapting and improving the model.

Literature

Dijk, T.J. van

Standaard Enquête Bevolking; bevolkingsonderzoek in verband met criminaliteitsbeleid ten behoeve van politie, bestuur en openbaar ministerie

The Hague, Ministry of Justice, DCP, 1991

Kruissink, M.

Measuring Police Performance; evaluation of the Dutch police. In: D. Dölling, Th. Feltes (eds.), Community policing - comparative aspects of community oriented police work

Holzkirchen, Felix-Verlag, 1993

Kruissink, M., G.J. Terlouw

Meten en vergelijken van politie-prestaties; eerste aanzet tot een 'sophisticated' model

Algemeen politieblad, vol. 141, no. 9, 1992, pp. 18-21

Terlouw G.J., C.J. Wiebrens

Bedrijven van opsporing; het meten van prestaties bij de politie Justitiële verkenningen, vol. 17, no. 3, 1991, pp. 40-70

Wiebrens C.J., M. Kruissink, G.J. Terlouw

Vergelijken van politieprestaties; aanzet tot de ontwikkeling van een output/effectrapportage van de politie

The Hague, Ministry of Justice, 1992

RDC K21