

## **Management and performance of territorial police structures: a conceptual framework based on indicators**

**Gabriel Nicolae UȚĂ**

Valahia University of Targoviste, Romania  
gabriel.uta@yahoo.com

**Abstract.** *The management of police structures involves the coordination of activities carried out in an environment characterized by uncertainty, operational pressure, legal responsibility and the use of limited public resources. Under these circumstances, the evaluation of management cannot be reduced to the number of interventions, sanctions or criminal acts recorded. The article proposes a multidimensional conceptual framework for identifying and formalizing indicators that could support the managerial evaluation of territorial police structures. The approach has a theoretical and exploratory character: no real data are processed, no concrete units are evaluated and no institutional rankings are formulated. Dimensions regarding the operational context, effectiveness, responsiveness, human resources, use of material and financial resources, quality of service, legitimacy and managerial capacity are discussed. The methodological part develops indicative formulas for expressing indicators, normalizing values, constituting sub-indices and building a synthetic managerial index. The numerical examples are exclusively hypothetical and have the role of clarifying the mathematical mechanism, not of producing empirical findings. The proposed framework can represent a basis for further research, provided that the indicators are validated, the data quality is ensured and the results are interpreted in relation to territorial and institutional specificities.*

**Keywords:** public management, police management, institutional performance, performance indicators, evidence-based management, legitimacy, composite index.

**JEL Classification:** H76, H83, M12, C81.

## 1. Introduction

The management of police structures is a field of research at the intersection of public management, public administration, security studies and organizational performance analysis. Its complexity results from the institutional specificity of police activity, which involves the exercise of public authority, the protection of citizens' rights and freedoms, the maintenance of public order, the prevention and fight against crime, as well as the management of situations characterized by uncertainty, temporal pressure and high levels of responsibility. In this context, the performance of police structures cannot be assessed solely by the number of activities carried out or by the volume of reported operational results. This must be analyzed in relation to the use of resources, the quality of the decision-making process, the coordination capacity, institutional adaptation, managerial stability and the effects produced on the safety of the community.

The economic, social and technological transformations of the last decades have significantly altered the environment in which law enforcement institutions operate. The diversification of forms of crime, the expansion of cybercrime, the mobility of the population, the intensification of the circulation of information and the increase in citizens' expectations have led to the emergence of additional requirements regarding the organization and management of police structures. At the same time, budgetary constraints, staff shortages, fluctuations in human resources and the need to modernize the technical infrastructure force managers to make informed decisions, set clear priorities and constantly monitor the relationship between the objectives set and the results achieved. Police management can no longer be limited, under these conditions, to the hierarchical coordination of activities, but must include tools for strategic planning, performance management, risk analysis and institutional impact assessment.

The particularities of police management are also determined by the strongly regulated and hierarchical character of the organization. Managerial decisions are adopted within a strict regulatory framework, and the autonomy of territorial structures is influenced by centrally established competences, internal procedures and the distribution of resources. The hierarchy ensures unity of command and coherence of interventions, but can lead, in certain situations, to organizational rigidity, delays in the decision-making process or difficulties in adapting to local particularities. From this perspective, one of the main challenges of police management is to strike a balance between institutional standardization and operational flexibility. Territorial structures must comply with general objectives and procedures, but at the same time they must respond to different local problems depending on the demographic, economic, geographical and criminal profile of each community.

Institutional performance appraisal is an essential component of results-oriented management. However, measuring performance in the field of policing entails numerous methodological difficulties. A low level of crime may be the result of effective preventive activity, but it may also be influenced by social, demographic or economic factors that are

not under the direct control of the police. Similarly, the increase in the number of crimes found may indicate a deterioration in the operational situation, but may also reflect a better ability to identify and document the facts. Therefore, the indicators used must be interpreted in context and analysed in an integrated manner. An evaluation based solely on a single indicator may produce incomplete conclusions, as the operational results must be correlated with the available resources, the volume of requests, the degree of staff load and the characteristics of the environment in which each structure operates.

Within this framework, the article aims to analyze the relationship between organizational management and the performance of territorial police structures, with a focus on how planning, coordination, use of resources and monitoring of results contribute to the achievement of institutional objectives. The research starts from the premise that the performance of a police structure is not determined exclusively by the number of police officers or the level of financial resources available, but also by the managerial capacity to organize these resources, set priorities, distribute responsibilities and adapt interventions to the needs of the community. Consequently, the managerial analysis must go beyond the strictly operative perspective and include dimensions related to efficiency, effectiveness, organizational quality, managerial continuity and adaptability.

The main objective of the research is to identify the managerial elements that can influence the performance of police structures and to highlight ways in which the management process can be analyzed in a coherent manner. In order to achieve this objective, several directions are pursued: presenting the particularities of management in public order institutions, examining the concept of performance in the police field, identifying relevant indicators for evaluating the activity and formulating an indicative mathematical model. At the same time, the study aims to highlight the need to move from a predominantly administrative and descriptive approach to an analytical one, in which managerial decisions can be supported by measurable information and interpretations made in relation to the institutional context.

Methodologically, the research combines the theoretical analysis of the concepts specific to public management and performance management with the mathematical formalization of a set of indicators. The theoretical component allows the clarification of the concepts of efficiency, effectiveness, leadership, planning, managerial control, legitimacy and institutional performance. The mathematical component has an exclusively demonstrative role and shows how certain dimensions could be expressed quantitatively in the case of valid data. The article does not process a database, does not estimate statistical models and does not evaluate a concrete police structure. Consequently, formulas must not be confused with empirical results and cannot substantiate, in the form presented, rankings or causal conclusions.

The importance of research derives from the need to develop evaluation mechanisms that provide a more complete picture of the functioning of police structures. The responsible

use of indicators can help identify imbalances, allocate resources more appropriately, prioritise and increase managerial transparency. Indicators should not be transformed into formal objectives pursued independently of operational reality, but should be used as diagnostic and decision-support tools. Their interpretation must be completed by analyzing the organizational context, qualitative evaluations, and examining external factors that may influence results.

## **2. Literature and identification of relevant indicators**

### **2.1. Performance Management in Police Institutions**

The evaluation of management within police structures must be analyzed in the broader context of the transformations produced in public management. The traditional administrative model has emphasised compliance with rules, the exercise of hierarchical authority, procedural control and the performance of tasks set at a higher level. From that perspective, the proper functioning of the institution was assessed mainly by the legality of its activities, organisational discipline and compliance with procedures. However, the development of performance-oriented public management has led to the extension of the evaluation criteria, by including effectiveness, efficiency, quality of services, use of resources and managerial responsibility. In the case of police structures, this change involves moving from simply reporting the activities carried out to examining the results and effects produced on community safety.

Performance management is the process by which institutional objectives are set, activities are planned, resources are distributed, and results are tracked and evaluated. In a police structure, this process is influenced by peculiarities that are not found to the same extent in other public organizations. Policing is carried out in an environment characterized by uncertainty, risk, time pressure, legal responsibility and the need to take quick decisions. In addition, the results of the activity cannot always be separated from the influence of economic, social, demographic and territorial factors. For this reason, police performance cannot be reduced to a single numerical value or to the evolution of a single phenomenon, such as the number of crimes recorded.

The literature highlights the fact that the police produce several categories of public value. A multidimensional normative framework includes reducing crime and victimization, reducing fear and strengthening personal safety, guaranteeing the security of public spaces, fair use of authority, efficient use of resources, quality of services and institutional legitimacy. This perspective leads to the need for a multidimensional evaluation framework, in which the operational results are analyzed together with the use of resources, the quality of the service and the relationship of the institution with the community. Subsequent studies show that the difficulty lies not only in defining indicators, but also in maintaining a balance between what is easy to measure and what is actually relevant to the organization's mission.

Managerial performance should not be confused with operative performance. Operative results refer to the activities and effects produced by the police structure, while managerial performance refers to how management sets priorities, organizes activity, distributes personnel, uses information, and manages available resources. A structure can achieve certain favorable results due to a relatively stable social context, even if managerial practices are insufficiently developed. Conversely, a structure operating in an environment characterised by high crime, intense mobility and staff shortages may perform apparently poorer, although its management makes efficient use of available resources. The evaluation must therefore differentiate the gross results from the results in relation to the actual operating conditions.

## **2.2. The multidimensional nature of police performance**

The activity of the police includes preventing and combating crime, maintaining public order, intervening at citizens' requests, conducting investigations, ensuring road safety, protecting vulnerable people and cooperating with other institutions. These responsibilities are different in terms of the resources required, their duration and their difficulty. Because of this, aggregating them into a single general indicator can lead to the loss of important information. An intervention at a routine event cannot be considered equivalent to managing a complex situation, even if both are recorded in the institution's records as completed activities.

The literature recommends differentiating between inputs, processes, immediate outcomes, and effects. Inputs include personnel, budget, working time, vehicles, equipment, and information infrastructure. The processes describe the actions carried out by the police, such as patrols, controls, interventions and investigations. Immediate results refer to the direct outputs of these activities, such as solving cases, carrying out interventions or carrying out preventive actions. Effects represent changes in safety, public order, citizens' trust or risks in the community. A coherent management system must follow all these levels, avoiding treating the volume of activity as the equivalent of the impact.

This delimitation is relevant because a high volume of activities does not automatically guarantee better results. The increase in the number of controls or sanctions may indicate an increase in activity, but does not directly demonstrate the reduction of risks. In certain situations, effective prevention of a problem can even reduce the volume of interventions needed. Expert analyses show that traditional measurement systems tend to place excessive emphasis on activities that are easy to count, while risk prevention and solving complex problems are more difficult to reflect through conventional indicators.

From a management perspective, a well-targeted preventive activity can be more valuable than a high number of reactive interventions. However, prevention is difficult to assess because it involves observing events that have not occurred. The absence of a negative event can be the result of police measures, but it can also be determined by social, economic

or behavioral changes. Preventive indicators must therefore be interpreted in relation to the nature of the risk, the target group, the implementation period and the characteristics of the area where the intervention has been taken.

### **2.3. Limitations of traditional indicators**

Among the traditional indicators used to describe police activity are the number of reported facts, crime rate, number of interventions, resolution rate, response time and number of sanctions applied. These indicators have the advantage of availability and relative ease of communication. However, each of them reflects only part of the activity and may be influenced by factors that are not under the direct control of management. Benchmarking analyses of measurement systems have shown that organisations continue to predominantly use traditional indicators, even when strategic objectives are formulated in a broader manner.

The number of reported facts may increase due to the intensification of the criminal phenomenon, but also as a result of a greater willingness of citizens to report incidents. Greater trust in the police, improved reporting systems or information campaigns can lead to an increase in the number of facts known to the institution. Under these circumstances, statistical growth should not automatically be interpreted as a deterioration in performance. Conversely, a low number of complaints does not necessarily confirm the existence of a high level of safety, as some categories of facts may remain unreported.

The resolution rate is relevant for the investigative activity, but is influenced by the complexity of the cases, the availability of evidence, the behavior of victims and witnesses, the duration of expertise and cooperation with other judicial institutions. Comparing structures that handle different types of cases can lead to erroneous conclusions. A structure that predominantly investigates complex facts may have a lower resolution rate than a structure that handles less difficult cases, without this difference demonstrating the existence of inferior management.

Response time is an important indicator for police services, especially in the case of urgent events. However, its value is influenced by distances, road infrastructure, crew distribution, degree of urbanization and number of simultaneous requests. Response time must be correlated with the urgency of the event and the quality of the intervention, as speed does not independently guarantee the adequacy of the response. An overall average can hide both situations resolved very quickly and delays in critical cases.

The number of sanctions, controls or actions may describe the intensity of the activity, but does not provide sufficient information about the effects produced. In addition, turning these metrics into rigid goals can lead to task orientation that generates quick numerical results. Thus, prevention, complex investigations, collaboration with the community and activities whose effects become visible in the longer term can be neglected. Indicators should support managerial judgment, not replace it.

## **2.4. Evidence-based management in policing**

Evidence-based management involves the systematic use of information to set priorities, choose interventions and evaluate results. In the police field, this concept is associated with the evidence-based policing model. This approach is structured around directing interventions to relevant problems, testing solutions, and tracking how they are implemented. Contemporary formulation emphasizes the need to combine the best available evidence with professional judgment and community values, preferences, and circumstances.

The application of this approach does not mean that managerial decisions should be based exclusively on numerical values. Data must be combined with professional experience, analysis of the local context, legal framework and risk assessment. Indicators can signal a problem, but they don't always explain its causes. For example, the increase in response time can be associated with staff shortages, inadequate distribution of crews, expansion of the territory covered or an increase in the number of requests. In order to establish a managerial measure, it is necessary to identify the relevant factor, not just to observe the change in the indicator.

In this article, evidence-based management is used as a basis for proposing categories of indicators that could support the evaluation of the activity. The article does not carry out the processing of data and does not aim at the statistical verification of the relationships between indicators. The aim is to identify information that could be relevant to the management of a police structure and explain their role in the managerial process. The proposed indicators are therefore conceptual in nature. They are not used to evaluate a real structure, to make a ranking or to draw conclusions about the effectiveness of a particular police unit.

## **2.5. Context and operational pressure indicators**

The first category that should be considered is the context indicators. They do not directly measure managerial performance, but are necessary for the correct interpretation of results. Among the relevant elements are the population served, the area of the territory, the population density, the share of the urban population, the mobility of the population, the number of localities and the economic or social particularities of the area.

Operational pressure can be described by the number of requests received, the number of interventions, the number of facts reported, the volume of cases in progress and the number of events requiring the mobilization of additional resources. These indicators show the level of demand addressed to the police structure. They are important because two structures with a similar number of police officers can operate in very different conditions if one of them handles a much higher volume of requests.

Contextual indicators should be used to avoid direct and simplified comparisons. The results of a territorial structure cannot be evaluated without reference to the characteristics

of the community and the level of operational pressure. A longer response time can be explained by the size of the territory and the dispersion of localities, as a high volume of reported facts can be associated with an urban center characterized by intense mobility and diversified economic activity.

## **2.6. Operational effectiveness indicators**

Effectiveness indicators should track the extent to which the structure is meeting its operational objectives. This category may include the rate of resolution of cases, the proportion of interventions properly completed, the reduction of some categories of risks, the evolution of serious events and the achievement of the objectives set by the action plans.

The resolution rate can provide information about investigative capacity, but it must be differentiated according to the type and complexity of the cases. Indicators on preventive activity should also track not only the number of actions carried out, but also their relevance to the problems identified. The mere organisation of activities does not demonstrate the achievement of effects, in the absence of a clear link between the measure and the risk pursued.

Effectiveness must be separated from the volume of work. The high number of actions, controls or interventions shows that resources have been used, but does not automatically confirm that the objective has been achieved. For this reason, quantitative indicators must be accompanied by evaluations of the quality of the intervention, the difficulty of the cases and the concrete outcome of the work.

## **2.7. Indicators on responsiveness**

The responsiveness reflects the structure's readiness to respond to citizens' requests and manage events within an appropriate timeframe. This category can include the response time, the proportion of urgent requests handled in the established intervals, the average duration of solving some works and the number of requests left uncompleted.

The response time must be interpreted differently depending on the nature of the event. Urgent requests should be analyzed separately from those that do not pose an immediate risk. Also, urban and rural environments cannot be assessed by identical standards without taking into account distances and crew availability.

The responsiveness is directly influenced by the managerial organization. The territorial distribution of personnel, shift planning, vehicle availability and the way requests are prioritized can produce important differences. From this perspective, indicators on the time and degree of completion of requests can provide useful information about the organization of the activity, but do not represent a complete measure of the quality of management.

## **2.8. Human resources indicators**

Human resources are one of the main components of institutional capacity. Relevant indicators may include job occupancy, number of vacancies, distribution of staff across

structures, average workload, staff turnover, absenteeism and participation in training activities.

The occupancy rate provides an overview of the availability of staff, but does not indicate whether the existing scheme is adapted to real needs. A structure can have a relatively high occupancy rate and, at the same time, face a shortage of personnel in certain departments. For this reason, the distribution of human resources is as important as the total number of employees.

Staff loading can highlight imbalances between the number of jobs and the number of available staff. However, the activities do not have the same level of complexity, and a simple average can hide important differences. Vocational training indicators may show interest in skills development, but the number of hours of training must be analysed together with the relevance of the programmes and the possibility of applying the knowledge acquired.

## **2.9. Indicators on the use of material and financial resources**

The management evaluation may also include indicators related to budget utilization, equipment availability, fleet status, access to technologies and costs of activities. These indicators are relevant for assessing management's ability to translate available resources into services and results.

The degree of use of the budget should not be confused with efficiency. Spending the funds in full does not automatically demonstrate that resources have been properly distributed. The analysis should follow the link between expenditure, institutional priorities and the results pursued. Also, the costs can vary depending on the type of interventions, the area of the territory and the technical needs of the structures.

Economic indicators should be used with caution in the field of policing. Cost reduction cannot be an independent objective when it affects the safety of staff, the quality of interventions or the protection of citizens. Efficiency should be understood as the rational use of resources to achieve objectives, not as the mechanical reduction of expenses.

Contemporary literature attaches major importance to the way citizens perceive the work of the police. Trust, satisfaction, procedural fairness and respectful treatment are dimensions that can influence the cooperation of the population with the institution. Research on procedural justice shows that legitimacy is associated not only with the outcome of an intervention, but also with the neutrality of the decision, the possibility of people to express their point of view, the respect shown in the interaction and the trust in the intentions of the authority.

This category can take into account the satisfaction of the people who have come into contact with the police, the number and nature of complaints, the proportion of confirmed complaints, the perception of the fairness of the treatment and the level of trust in the institution. This information can be obtained through questionnaires, interviews or the

analysis of complaints. The methodological analyses highlight the need to use multiple measures to assess the public's perspective and the risk of misinterpretation of general questions on trust or satisfaction.

The number of complaints should not be interpreted in isolation. An increase can be determined by real problems, but also by a better knowledge of the referral procedures. Similarly, a small number of complaints does not automatically demonstrate the existence of a high quality of service. The data must be analysed in relation to the number of interactions, the subject of the complaints and the result of their verification.

In order to evaluate the management itself, the stability of the management, the continuity of strategies, the degree of achievement of objectives, the existence of monitoring mechanisms, the frequency of internal evaluations, the ability to correct deviations and the level of cooperation between departments could be analyzed.

Managerial stability can favor the continuity of projects and the maintenance of a strategic direction, but it should not automatically be interpreted as a favorable outcome. Stable leadership can maintain effective practices, but it can also preserve procedures that no longer correspond to current problems. Stability must therefore be analysed together with adaptability and institutional outcomes.

The degree of achievement of the objectives can only be a useful indicator if the objectives are clearly defined, measurable and realistic. Purely quantitative objectives can encourage a formal orientation towards achieving values, without solving real problems. The evaluation must aim at both the achievement of the objective and its relevance to the needs of the community.

The ability to adapt can be seen in the speed with which management identifies problems, redistributes resources, modifies plans and introduces corrective measures. This dimension is difficult to measure exclusively numerically, which is why it should be completed by analyzing the management documents, procedures and decisions adopted.

The literature analyzed supports the use of a balanced system, which combines operational, organizational, economic and social indicators. There is no single indicator by which the performance of a police structure can be fully assessed. There is also no universal framework that can be applied without adaptation to all structures, regardless of size, attributions and territorial context. The application of the Balanced Scorecard in police organizations aimed precisely to extend the evaluation beyond immediate operational results and to connect indicators with strategic objectives.

A balanced system could include the following dimensions: context and operational pressure, effectiveness of activities, responsiveness, human resources situation, use of financial resources, quality of service, legitimacy and managerial capacity. Indicators should be interpreted together, as the improvement of one dimension can be associated with the deterioration of another.

For example, reducing response time may mean focusing crews on urgent interventions and decreasing the resources available for prevention. Increasing the resolution rate can be achieved by prioritizing simpler cases, while complex investigations receive fewer resources. An evaluation based solely on one of these indicators could reward managerial behavior that does not produce the most important effects for the community.

Indicators should be used as diagnostic tools, not as substitutes for managerial analysis. Their role is to flag differences, trends, and possible issues that require further examination. The interpretation must take into account the context, the quality of the data and the limits of each measure. Therefore, the present research proposes a conceptual framework of indicators, without calculating or testing them on the basis of actual data.

### **3. Conceptual model and mathematical formalization of the proposed indicators**

#### **3.1. Methodological status of formalization**

The formalization proposed in this section has a conceptual and demonstrative role. It does not automatically transform police management into a completely quantifiable phenomenon and does not assume that the differences between territorial structures can be explained by a single numerical value. The goal is narrower: to define a coherent architecture through which administrative, operational and organizational information could be ordered, compared and interpreted if there were complete, stable and validated data. In this sense, mathematics is used as a language for disciplining concepts, not as a substitute for professional judgment. Each component must have a clear institutional meaning, an identifiable data source, and a rule of interpretation that distinguishes between context, managerial effort, result, and effect.

The unit of analysis is represented by the territorial structure observed in a constant reference period, which can be monthly, quarterly or annual. The choice of period must avoid mixing phenomena with very different rhythms: the response time may vary from one day to the next, the occupation of positions changes more slowly, and the effects on public trust may become visible only at longer intervals. For this reason, the framework must be regarded as a succession of analytical levels and not as a mechanically executed calculation. Raw values are first defined, then contextualised, favourably or unfavourably oriented, brought on a common scale and only then aggregated. Any empirical application should maintain this order in order to avoid turning purely administrative differences into apparent differences in performance.

An important consequence of this positioning is the separation between mathematical validity and managerial validity. An expression can be impeccably calculated and still give a poor picture of the phenomenon if the selected variables do not capture the mission of the institution, if the denominator is not comparable or if the indicator reacts to external factors. In applied research, validation should include at least three levels: checking the definitions

together with practitioners, testing the statistical properties of the series, and comparing the results with qualitative information about the decisions made during the periods analyzed. This triangulation is especially necessary in law enforcement organizations, where results are produced through the interaction between legal rules, resources, organizational behavior, and community characteristics. Formalization is useful precisely because it makes the hypotheses visible, but it cannot decide on its own whether those hypotheses are correct.

In addition, any outcome should be accompanied by a description of its uncertainty. In the case of administrative data, the uncertainty stems not only from statistical error, but also from changes in procedures, differences in recording and conceptual limits of the indicator. Reporting these limits does not weaken the model; on the contrary, it prevents its use beyond what the information can support and preserves the distinction between measurement, interpretation and managerial decision.

### **3.2. Indicator architecture and working notation**

The framework uses a narrow set of variable families so that the model remains intelligible. The context data include the population served, the area of the territory, the density, the degree of urbanization, the mobility and the particularities of the network of localities. Demand for services data includes referrals, interventions, causes or works under management and events that require the simultaneous mobilisation of several resources. Internal capacity is described by planned and employed personnel, personnel with operational attributions, technical availability of vehicles and equipment, financial resources and actual working time. The results are treated separately from these entries and include the resolution of the works, the meeting of deadlines, the response to requests, the achievement of objectives and the quality of interaction with the public.

Notation should not become more important than definitions. Two databases can use the same label for "case solved", but they can apply different rules regarding the time of registration, stock from previous periods or the reopening of works. Before any calculation, each variable must be accompanied by a technical sheet specifying the numerator, denominator, unit of measurement, source, periodicity, reporting officer and rules for dealing with missing values. The same discipline is required for qualitative variables converted into scores. Managerial stability, internal cooperation or the relevance of professional training cannot be credibly introduced into a model just by assigning a grade; they require observable criteria and reproducible assessment procedures.

To avoid overloading the model, each variable should only be introduced if it brings distinct information and can be interpreted managerially. Highly correlated indicators can describe the same reality in different forms and lead to the implicit doubling of the importance of a dimension. For example, the number of requests, the number of interventions and the volume of calls can evolve almost identically in a given structure, in which case aggregating them without verification would overestimate the operational pressure.

Conversely, two seemingly similar indicators can capture different phenomena if one describes the inputs and the other the completion of activities. Therefore, a conceptual mapping stage is recommended before the final selection, in which each indicator is linked to the managerial question to which it answers. This step reduces the risk of building a data-rich system that is poor in decision-making significance.

An initial inventory of the data should also include the level at which each variable can be observed: structure, compartment, locality, type of event or category of cause. Premature aggregation can hide the focus of a problem in a tight area, while excessive granularity can lead to instability and data protection difficulties. The choice of the level of analysis must, therefore, be linked to the managerial decision pursued. For the distribution of resources, territorial and temporal granularity is useful, while quarterly or annual values may be sufficient for strategic evaluation. The coherence between the question, the unit of analysis and the indicator is a basic condition for interpretation.

### **3.3. Context and operational pressure**

Operational pressure must be treated as a contextual dimension, not as a negative outcome attributed to management. A structure that serves a larger population, a larger territory or an environment with intense mobility may use more resources and record higher response times without this automatically indicating poorer management. From this perspective, the absolute number of notified facts, requests or interventions is not sufficient. The information becomes relevant when it is related to the population, area, operative staff and complexity of activities. At the same time, the comparison must avoid the assumption that all requests have the same weight. A simple event and an intervention involving several crews, trips and subsequent activities cannot be treated as identical work units.

For the conceptual model, the pressure is built from four nuclei: the phenomenon known to the institution in relation to the population, the demand for services in relation to the population, the spatial density of the interventions and the workload in relation to the operative staff. The last component can integrate the causes in progress, the outstanding stock and the urgent requests through complexity coefficients. Territorial and mobility factors are used as contextual amplifiers, because the same numerical load can involve different effort in a compact municipality compared to a county with dispersed localities. The following phrase summarizes these relationships in a single block, without assuming that the weights or coefficients are already empirically validated.

Contextualization can also be developed by grouping structures into comparable categories, without turning the grouping into a ranking. Units with high population density, intense mobility and high volume of nocturnal events can be analyzed separately from predominantly rural structures, where distances and dispersion of localities produce other constraints. Such stratification allows for the use of more realistic landmarks and reduces the risk of a structure appearing to be performing or underperforming just because it is compared to a unit in a radically different environment. At the same time, the pressure must

be followed in the dynamics. A moderate but persistent increase in requests may be more managerially important than an isolated peak produced by an exceptional event. Because of this, the analysis of successive periods and the identification of trends must complement any aggregate value calculated for a single interval.

It is also recommended to identify exceptional events that can temporarily alter the profile of a structure: large-scale events, sporting events, severe weather phenomena, tourist periods or special operations. These episodes should not be removed from the data, as they are part of the actual activity, but they should be tagged. Thus, the researcher can separate the structural, persistent pressure from the episodic pressure and observe whether the organization returns to its usual level after the event has ceased.

$$\Pi_{it} = 100 \left[ \alpha_1 \frac{F_{it}/P_{it}}{q_F} + \alpha_2 \frac{S_{it}/P_{it}}{q_S} + \alpha_3 \frac{I_{it}/A_i}{q_I} + \alpha_4 \frac{C_{it} + \rho B_{it} + \eta U_{it}}{E_{it}^{op} q_C} \right] (1 + \lambda_1 \tilde{D}_{it} + \lambda_2 \tilde{M}_{it}), \quad \sum_{r=1}^4 \alpha_r = 1, \quad \alpha_r, \rho, \eta, \lambda_1, \lambda_2 \geq 0.$$

In this expression, the weighting coefficients control the contribution of each source of pressure, and the territorial parameters adjust the interpretation without turning the context into a managerial penalty. Benchmarks should be established prior to empirical application, on a comparable and documented basis.

### 3.4. Operational effectiveness and achievement of objectives

Effectiveness describes the extent to which the work produces the desired results, but must be separated from the mere volume of work. A high resolution rate can have different meanings depending on the complexity of the case portfolio, and exceeding a quantitative target is not always a positive outcome. If the objective is defined by the number of controls, sanctions or actions, the mechanical tracking of the target can stimulate the production of activity that is easy to report, without the guarantee of a relevant effect for the community. Therefore, the evaluation of effectiveness must combine the final result with the difficulty of the activities and the relevance of the objective. In an applied system, the categories of cases should be grouped according to stable criteria, and the weights of complexity should be substantiated by average working time, resources involved or documented professional evaluations.

Another necessary distinction is that between prevention and reaction. The number of preventive actions cannot alone demonstrate that the risk sought has been reduced, but the structure of the activity can indicate whether the management reserves resources for interventions before events occur. The analysis must follow the link between the identified problem, the population or area targeted, the selected measure and the subsequent development. Routine preventive activities, carried out without a risk-based justification, should not be given the same importance as measures targeting documented problems. Consequently, effectiveness is treated as a combination of the completion of activities, the

achievement of the relevant objectives, the complexity of the portfolio and the ability to produce effects without the incentive of artificial reporting.

The evaluation of the objectives must also include the quality of the process by which they were set. An easy-to-achieve goal does not demonstrate performance, just as a very ambitious goal, formulated without regard to resources and context, can artificially produce a low rate of achievement. In practice, objectives can be classified into mandatory outcomes, improvement targets and exploratory outcomes, with each category having a different interpretation. For mandatory objectives, failure to achieve it may signal a problem of capacity or coordination; for improvement targets, the direction and pace of progress matters; For exploratory goals, organizational learning may be relevant even if the numerical target is not fully achieved. This distinction limits the pressure to transform all activities into achieved/unachieved binary values and preserves the link between measurement and the managerial sense of the goal.

Effectiveness can also be tracked by the persistence of the result. An improvement observed in a single period is not sufficient to conclude that a managerial measure has produced a stable effect. It is preferable to examine several periods before and after the intervention, paying attention to external changes that may explain the evolution. In this way, the framework moves closer to an evidence-based assessment and reduces the risk of attributing conjunctural results to a managerial decision.

### **3.5. Responsiveness and management of the stock of works**

Responsiveness is one of the most visible dimensions of the police service, but also one of the most context-sensitive. The average response time is intuitive, but can be distorted by a small number of extreme events. The median, the distribution by priority classes and the proportion of requests within a reference range provide a more stable picture. Urgent events should be analysed separately and reference intervals should be determined according to the nature of the request and territorial conditions. An aggregate indicator that combines critical cases and non-urgent requests can hide exactly the situations on which management needs to intervene. Also, a quick response is not enough if the intervention is incomplete, inadequate or generates repeated replays of the same problem.

Inventory management complements the response time analysis. A structure can respond quickly to new events, but it can accumulate papers, petitions, or backlogs. For management, both the size of the stock and its age, the share of cases with overruns and the concentration of arrears in certain compartments are important. The stock should be related to the entries in the period and the available resources to avoid simplistic interpretations. A higher volume may reflect a temporary increase in demand, a problem with staff distribution, excessively cumbersome procedures or an increase in complexity. The indicator becomes useful only if it is accompanied by a managerial analysis that identifies the source of the accumulation and corrective measures.

Time analysis can be improved by separating the internal stages of the response. The total time between notification and intervention includes, depending on the system, the interval of retrieval and classification, the time until the allocation of the crew, the actual movement and any delays caused by the unavailability of resources. If all these components are aggregated into a single average, management cannot identify the exact source of the problem. A performance management system should allow for the breakdown of the delay and track separately cases where the standard has been exceeded. The same logic applies to the stock of works: the age, the cause of the exceedance, and the number of transfers between departments can be more useful than the total value. Thus, temporal indicators become tools for diagnosing processes, not simple speed measurements.

For high-priority events, the complete distribution of times is more important than the average. Very fast case rates, near-threshold cases, and severe delays may indicate different problems. A management may have an acceptable average but a small number of critical delays that require immediate analysis. For this reason, reporting should also preserve information about the tail of the distribution, not just a central value.

### **3.6. Human resources and integrated operational capacity**

Human resources should not be analyzed exclusively by the degree of occupation of positions. The approved scheme may itself be insufficient for the actual volume of activity, and the internal distribution may produce a deficit in critical compartments even when the overall employment rate is high. Therefore, the evaluation must follow the personnel actually available, the operative staff on shifts, the differentiated load by types of activities, the fluctuation, absenteeism and access to training. Fluctuation requires separating retirements, transfers, promotions, and unplanned departures, as the managerial effects are different. Professional training must be assessed by relevance, completion and applicability, not by the gross number of course hours.

However, operational capacity results from the interaction of staff with other resources and with the results achieved. A structure with full staff, but with unavailable vehicles or with a large portfolio of outstanding works, cannot be adequately described by the occupancy indicator. For this reason, the second mathematical block brings together five components: complexity-weighted resolution, compliance with priority response intervals, effective occupancy, fleet availability and an inversely oriented component on confirmed complaints. The score is not proposed as a final measure, but as an example of how resources and outcomes can be connected without confusing them.

In the evaluation of human resources, the structure of competences must also be taken into account, not just the numerical number of personnel. Two departments with the same number of positions occupied may have different capacities if one has experienced staff, specialised qualifications and continuity, and the other is in a period of integration of new employees. An applied approach could follow the distribution of seniority, critical skills, reliance on a small number of specialists and the existence of succession plans for sensitive

functions. This information should not be used for the individual assessment of police officers, but for the identification of organisational vulnerabilities. From a managerial perspective, capacity means the existence of the right person, with the right competence, in the right place and at the right time; The mere administrative occupation of a position does not guarantee that this condition will be satisfied.

Exchange planning is a distinct link between workforce and performance. The total availability of staff in a month may be satisfactory, but the distribution by hours and days may not be aligned with the peak moments of requests. Comparing the demand curve with the actual presence makes it possible to identify imbalances that do not appear in the annual indicators. This perspective transforms the human resource from an administrative stock into a capacity that must be allocated temporally according to the risk.

$$K_{it} = 100 \left[ \beta_1 \frac{\sum_{c=1}^C \kappa_c Sol_{cit}}{\sum_{c=1}^C \kappa_c Reg_{cit}} + \beta_2 \frac{\sum_{j=1}^J \omega_j 1(t_{jit} \leq \tau_j)}{\sum_{j=1}^J \omega_j} + \beta_3 \frac{E_{it}^{occ}}{E_{it}^{plan}} + \beta_4 \frac{V_{it}^{disp}}{V_{it}^{alloc}} + \beta_5 \left( 1 - \frac{R_{it}^{conf}}{\max(R_{it}^{ver}, 1)} \right) \right], \quad \sum_{r=1}^5 \beta_r = 1, \quad \beta_r \geq 0.$$

The integrated score is deliberately illustrative. The components can be kept separately in the reporting, and the aggregation should only be used after checking the quality of the data and the managerial sense of the weights. A high value is not a substitute for case analysis and cannot demonstrate managerial causality.

### 3.7. Material and financial resources

The analysis of material and financial resources must start from the idea that budget execution is not synonymous with efficiency. Full use of the budget may reflect good planning, but it can also indicate spending funds to avoid losing them at the end of the period. Conversely, sub-execution can be associated with savings, delays in procurement, cancellation of procedures or unavailability of necessary products. The budget indicator must be correlated with the financed objectives, the timing of the acquisitions and the results pursued. Unit costs are only useful when there is a credible accounting delineation and when the activities compared are sufficiently similar. The reduction of the average cost cannot become an autonomous target if it affects the safety of the staff, the legality of the intervention or the quality of the service.

Technical availability provides a more direct link to daily operation. For vehicles, it is preferable to measure the days of availability or the hours in which the resource can actually be used, rather than the mere existence in the inventory. For equipment, the need must be established by categories and functions, and the surplus in one category must not compensate for the lack of essential equipment in another. In a managerial analysis, resources should be tracked through a table that combines allocation, availability, utilization, maintenance cost, and effect on activities. Such an approach makes it possible

to differentiate between a lack of resources and the inefficient use of already existing resources.

Another relevant element is the life cycle of resources. The purchase of a vehicle or equipment produces an apparent increase in capacity, but the managerial value depends on availability, maintenance, compatibility with the other systems and subsequent costs. Investments should be analysed in relation to the period during which the resource can actually be used and the activities it supports. In addition, differences between the approved budget and the resources actually accessible may arise from procurement procedures, delivery times or administrative restrictions that are not fully controlled by the local management. To avoid erroneous conclusions, the model must distinguish between allocation, hiring, payment, and actual use. This separation is essential to identify whether a problem stems from a lack of funding, a delay in implementation, or the misuse of an available resource.

For major investments, the assessment may include a pre- and post-implementation phase. It is not enough to notice that a resource has been purchased; It should be checked whether the downtime has been reduced, whether interventions have become faster or whether certain activities can be carried out with less effort. In the absence of such a link, expenditure remains an input indicator. Managerial analysis aims to transform the input into operational capacity and then into the result, while maintaining the separation between these levels so as not to automatically attribute effects to an acquisition.

### **3.8. Quality of service, integrity and legitimacy**

The quality of the service cannot be reduced to the number of complaints. The absolute number is dependent on the volume of interactions with the public, the accessibility of the notification procedures and the willingness of citizens to complain. More informative is the reporting of complaints to interactions and the separation of those confirmed from those not confirmed or still under verification. Even so, the indicator must be interpreted together with the nature of the complaint, the seriousness of the aspect found and the compartment involved. A communication problem and an integrity breach cannot carry the same weight. For management, the usefulness of the indicator derives from the identification of repetitive patterns and the ability to introduce corrective measures, not from simply reducing the number reported.

Satisfaction, procedural fairness and institutional trust require distinct measuring tools. Questionnaires must use multiple items, have stable sampling rules, and allow comparison over time. Procedural justice can be observed through the person's ability to express his or her point of view, the neutrality of the decision, the respect shown and the perception of the authority's intentions. Trust can include the competence, impartiality, integrity, predictability, and availability of the institution. These dimensions should not be aggregated before checking the internal consistency and how respondents understand the questions. In a mature managerial system, quantitative results are complemented by

qualitative analysis of interactions and examination of cases indicating institutional vulnerabilities.

The measurement of public perception must also distinguish between the general population and people who have had recent contact with the police. The former can evaluate the institution based on the public image, indirect experiences and information from the media space, while the latter can appreciate concrete aspects of the interaction. Both perspectives are relevant, but they answer different questions. For management, it is useful to track the difference between overall trust and satisfaction after contact, as well as to identify groups or areas where perceptions deviate significantly. The results should be interpreted with caution when the questionnaire response rate is low or when the sample excludes important categories. A legitimacy score acquires decisional value only if the collection methodology is stable enough to separate a real change from a simple change in the composition of respondents.

Complaints and questionnaires can be completed by analyzing recurrence. If the same types of dissatisfaction occur repeatedly in a compartment or in a certain category of interactions, the managerial signal is stronger than an isolated increase in the total number. Recurrence may indicate a procedural, training, supervision or communication problem. Consequently, the model must allow the transition from the overall score to the concrete operational cause that can be corrected.

### **3.9. Management organization and quality**

Indicators concerning management itself must describe processes over which management has an identifiable influence. Leadership team stability is relevant because frequent changes can disrupt projects, alter priorities, and reduce continuity of accountability. However, stability is not a positive value in itself: stable leadership can maintain both effective practices and persistent dysfunction. Therefore, the duration of mandates must be correlated with the achievement of objectives, adaptation to new problems and implementation of measures resulting from audit or control. More important than mere continuity is the ability to maintain a strategic direction and, simultaneously, to quickly correct decisions that do not produce the expected results.

Managerial adaptability can be analyzed by the sequence between detecting the problem, interpreting the information, deciding to redistribute resources, implementing the measure and verifying the effect. A high-performance managerial structure is not the one that permanently modifies the plans, but the one that can demonstrate that the change was justified by the data and that the result was followed. Compliance with deadlines for managerial measures, implementation of control recommendations and cooperation between departments provide additional clues. These elements are difficult to synthesize exclusively numerically and must be documented through decisions, plans, minutes, evaluation reports and other organizational traces. The role of formalization here is to organize the evidence, not to eliminate institutional analysis.

The quality of management can also be observed by the way information circulates within the organization. Data-driven decisions require channels through which operational information reaches the management level, is analyzed and returns to the structure in the form of clear priorities. Information bottlenecks can cause delays even when resources are sufficient. Therefore, empirical research could track the frequency of operative meetings, the existence of dashboards, the time between the identification of a problem and the issuance of a measure, as well as the mechanisms by which the effect of the decision is subsequently verified. It is not the number of meetings that is relevant in itself, but the link between information, decision and action. This circuit provides a closer perspective to management than purely operational indicators and can explain why structures with similar resources achieve different results in comparable situations.

A complementary element is the quality of delegation. Excessive concentration of decisions at the top of the structure can produce coherence, but it can slow down the reaction to local problems. Delegation without control mechanisms can instead lead to uneven practices. A managerial evaluation can track what types of decisions are decentralized, what information is needed for approval, and how long the decision-making circuit takes. The goal is not to maximize delegation, but to find a balance between speed, responsibility, and control. This balance can explain differences in adaptability even between structures operating under the same regulatory framework.

### **3.10. Data quality and comparability rules**

Systematic data quality control is required before standardisation. It must aim for completeness, consistency of definitions, stability of series over time, duplications, procedural changes and impossible values. A sudden drop in an indicator can reflect a real improvement, but also a change in the way it is recorded. Similarly, the introduction of a new IT system may temporarily increase the number of complaints or reduce administrative delays without operational management having changed to the same extent. Metadata should be kept along with values, and serial breaks should be marked explicitly. Comparability between structures is credible only when the same rules for data production are applied to all analyzed units.

Missing values should not be automatically replaced with zero, because zero is a substantive information different from the absence of information. Imputation may be justified for certain series, but must be reported and tested by sensitivity analyses. Extreme values must be investigated before any statistical limitation. Some may stem from errors, but others may signal a genuine operational situation, such as a major event, severe staff shortages, or a period of exceptional demands. Removing them automatically would precisely reduce the model's ability to identify important pressures. Therefore, data processing must be documented as part of the methodology, not treated as a neutral technical step.

Data traceability is just as important as its accuracy. For each value used in an indicator, it must be possible to identify the initial source, the transformation rule and the time of extraction. If the data is corrected later, the version used in the calculation must be kept for the reproduction of the result. A register of methodological changes should indicate changes in definition, the introduction of new categories and periods when IT systems have been unavailable. In the absence of this documentation, a difference between two years may reflect a technical change and not a managerial one. For research, reproducibility is an essential condition: another researcher should be able to reconstruct the indicator starting from the same data and rules. This requirement reduces the reliance on informal interpretations and turns the measurement system into an auditable tool.

Data control should include automated checks and human checks. Automatic rules can detect impossible values, unusual jumps, or incompatibilities between fields, while professional review can explain whether the deviation is an error or a real event. The separation of the two stages prevents both the mechanical acceptance of wrong data and the discretionary correction of unusual values. Each data intervention must be recorded in order for the final series to remain auditable.

### **3.11. Normalisation of indicators**

The indicators used in the model have incompatible units: minutes, percentages, number of events, costs, days, questionnaire scores, or population-related rates. For aggregation it is necessary to transform them on a common scale, but the choice of method influences the result. Favourable indicators, such as job occupancy or the proportion of interventions carried out within the reference interval, should be targeted so that a higher value corresponds to a better score. Unfavorable indicators, such as response time, outstanding stock or confirmed complaint rate, must be reversed. For indicators with an optimal target, deviation in either direction can be unfavorable. Normalisation should be established before the results are observed in order to reduce the risk of adjusting the rules according to the ranking obtained.

In the case of comparisons between structures, the minimum and maximum limits may come from the observed distribution, from institutional standards or from ex-ante defined thresholds. The use of observed values produces relative scores: one structure can get a high score just because the others perform worse. The use of external standards allows for more stable interpretation over time, but presupposes the existence of justified reference values. Neither option is neutral. Therefore, the methodological report must specify for each indicator the meaning, benchmark and treatment of extreme values. The third mathematical block concentrates in a single expression the three normalization orientations, the formation of subscripts and the geometric aggregation of the six dimensions.

The choice of the reference period for normalisation must also be carefully determined. If the limits are recalculated annually, the scores can remain apparently stable even when all

structures improve or deteriorate simultaneously, because the relative position is preserved. If the limits are fixed for several years, the temporal evolution is easier to follow, but the values can go out of the initial range when the system changes substantially. One solution is to report a relative score and a raw indicator at the same time or related to a stable target. In this way, management can observe both the position vis-à-vis other structures and the absolute progress. Normalization must therefore be seen as a stage of translation of information, not as a transformation that eliminates the need to see the initial values.

For managerial communication, normalized scores must always be accompanied by their meaning. A score of 70 doesn't have the same meaning if it comes from a job occupancy rate, response time, or satisfaction questionnaire. The scoreboard shall allow for a return to the gross value and the baseline. This transparency reduces the risk of scores being treated as abstract units, detached from the institutional phenomenon they represent.

$$N_{kit}^P = \frac{x_{kit} - x_k^{\min}}{x_k^{\max} - x_k^{\min}}, \quad N_{kit}^N = \frac{x_k^{\max} - x_{kit}}{x_k^{\max} - x_k^{\min}}, \quad N_{kit}^T = \max\left(0, 1 - \frac{|x_{kit} - T_k|}{\delta_k}\right).$$

$$D_{dit} = \sum_{k \in d} w_{kd} N_{kit}, \quad \sum_{k \in d} w_{kd} = 1, \quad IM_{it} = 100 \prod_{d=1}^6 \left(\frac{D_{dit}}{100}\right)^{v_d}, \quad \sum_{d=1}^6 v_d = 1.$$

The relationship focuses the entire process: orientation of indicators, transformation on a common scale, aggregation into sub-indices and obtaining a synthetic index. Operational pressure remains outside the final product and is used to contextualize the score, not to automatically diminish it.

### 3.12. Construction of sub-indices and synthetic managerial index

Aggregation must be carried out in two steps. First, dimensional sub-indices are constructed for operational effectiveness, response capacity, human resources, material and financial resources, quality of service and legitimacy, respectively organization and managerial capacity. This structure prevents numerous indicators in a category from automatically dominating the final result. Operational pressure is kept separate and used for interpretation, not as a direct penalty. Thus, a structure that operates in a difficult environment can be compared with other units without turning the external context into a managerial deficiency. The sub-indices should be published together with the overall index, as the aggregate value does not show where the strengths and vulnerabilities lie.

In the second step, the subscripts can be aggregated arithmetically or geometrically. The arithmetic mean allows for greater compensation between dimensions: a high score on effectiveness can compensate for a poor score on legitimacy. The geometric mean reduces this compensation and penalizes imbalances more sharply. For areas considered non-compensable, such as legality or integrity, a minimum threshold may be introduced, but the threshold must be normatively justified and empirically tested. The index should not be used as an automatic classification or sanctioning mechanism. Its main function is

diagnostic: it signals combinations of results that require investigation, facilitates the tracking of evolution over time and allows the analysis of how managerial changes are reflected in several dimensions simultaneously.

Sub-indices can also be used to build vulnerability profiles. Instead of a single order, structures can be grouped according to the configuration of scores: high pressure and adequate capacity, high pressure and insufficient resources, good effectiveness but low legitimacy, available resources but poor managerial processes. These profiles are more useful for decision-making because they suggest different types of intervention. Some issues require redeployment of staff, others review of procedures, training, investment or interventions on the relationship with the community. The general index can serve as an initial signal, but managerial action must be based on the internal structure of the score. In this sense, the model does not aim to simplify reality down to a single digit, but to organize it in a form that allows the transition from finding to diagnosis and then to corrective measure.

In longitudinal reporting, it is useful to keep the same subscript structure for a sufficient period. Frequent changes in composition can make it impossible to distinguish between changing performance and changing methodology. When an indicator needs to be replaced, the old and new series can be calculated in parallel for a transitional period. This overlap allows the estimation of the methodological effect and maintains the continuity of the managerial interpretation.

### **3.13. Weighting and governance of the model**

Weights are one of the most sensitive choices of a composite index. Equal weights are easy to explain and reduce the impression of artificial precision, but it does not mean that all components have the same institutional importance. An alternative is structured expert consultation, in which the importance of each dimension is independently assessed and then compared. Pair comparisons or other prioritization procedures can also be used, but technical complexity does not automatically guarantee a better decision. Transparency is essential: who set the weights, based on what criteria, with what degree of agreement and what effect they have on the results. If experts express widely dispersed opinions, simple media can hide the lack of consensus.

The governance of the model must also provide for the periodic review of the weights. The mission, risks and availability of data can change, and a set of weights fixed for a period of time can become inappropriate later. However, the revision should not be carried out frequently or opportunistically, as this would compromise temporal comparability. One solution is to establish a review cycle and keep both the series calculated with the original weights and a recalibrated series when the methodology changes. In this way, the system remains adaptable without losing analytical continuity. Decisions on weights, thresholds and definitions should be documented in a methodological protocol separate from the performance reporting itself.

An additional principle is the separation between those who manage the data and those who approve the evaluation rules. If the same structure can change the definitions, weights and values reported, the risk of the system being adjusted to produce a favourable result increases. Methodological governance should provide for distinct responsibilities, documentation of changes and an approval procedure before the new rules enter into force. For research uses, all versions of the methodology must be archived. For internal managerial uses, it may be useful to have a mixed group with operational, legal, statistical and control skills. This diversity reduces the risk that indicators are designed exclusively from the perspective of ease of calculation or, conversely, exclusively from the perspective of operational experience without verification of technical properties.

Before operational use, the model should be tested in a pilot phase. Piloting makes it possible to observe indicators that cannot be fed regularly, definitions interpreted differently and situations in which weights produce unexpected effects. The pilot period should not be used for the formal evaluation of structures, but for the correction of the methodology. After stabilizing the rules, the final version must be frozen for the agreed reporting interval. This discipline limits ad-hoc changes and gives users the opportunity to understand how each result is produced.

#### **3.14. Sensitivity analysis and robustness verification**

A composite index is only credible if its results do not change disproportionately to small and plausible variations in methodological assumptions. The sensitivity analysis follows exactly this aspect. Alternative sets of weights, different limits for normalization, alternative treatments of extreme values and different aggregation methods can be tested. If the relative position or conclusion on a structure changes sharply between scenarios, the outcome should be presented as unstable. Conversely, if the values remain close, the index can be used with a higher degree of confidence for descriptive analysis. Sensitivity should not be reduced to a single isolated change; a set of coherent scenarios is preferable, each representing a possible methodological interpretation.

The last mathematical block expresses this logic by recalculating the index for several weighting schemes, measuring the difference from the baseline scenario, and constructing a range of results. Included in the same area is an exclusively demonstrative numerical example with six hypothetical dimensional scores. The values do not belong to any real structure, and the result should not be interpreted as an institutional evaluation. The example is only intended to show that an explicit weighting scheme can be traced back to the aggregate value and that the change in assumptions can be quantified transparently.

Robustness can also be assessed by the controlled exclusion of one component at a time. If the removal of a single indicator radically alters the result, that variable may dominate the model or the other components may provide insufficient information. Another check is to recalculate the score over different periods, to see if the results depend on a particular range. For questionnaire-based indicators, sensitivity may include scenarios on response rate and

sample structure. For administrative data, alternative rules for handling stocks and missing values can be tested. The goal is not to choose the scenario that produces the preferred outcome, but to identify areas where the conclusion is fragile. A robust model is not one that produces exactly the same value under any conditions, but one whose main interpretation remains consistent in the face of reasonable methodological assumptions.

The results of the sensitivity analysis can be communicated through stability ranges and classifications. An index that varies very little between scenarios can be considered robustly descriptive, while a result with high variation must be accompanied by a methodological caveat. It is important that this information is not hidden in technical annexes, as it directly influences how managers can use the score. A seemingly accurate but very weight-sensitive value is less informative than a transparent range that shows the limits of the conclusion. In this way, robustness analysis becomes part of the interpretation and not just a separate statistical check.

$$IM_{it}^{(s)} = 100 \prod_{d=1}^6 \left( \frac{D_{dit}}{100} \right)^{v_d^{(s)}}, \quad \Delta_{it}^{(s)} = IM_{it}^{(s)} - IM_{it}^{(0)}, \quad [IM_{it}^{\min}, IM_{it}^{\max}]$$

$$= [\min_s IM_{it}^{(s)}, \max_s IM_{it}^{(s)}].$$

$$D = (78, 72, 81, 69, 75, 77), \quad v = (0.18, 0.17, 0.16, 0.15, 0.18, 0.16), \quad IM^{(0)} \approx 75.31.$$

The numerical example only shows the calculation mechanism. In real research, the range between the minimum and maximum scenario should be related together with the base value, and any unstable conclusion to reasonable changes in weights should be made with reservations.

### 3.15. Managerial interpretation of results

The final result should be read along with the dimensional profile and operational pressure. Two structures can achieve the same overall index through very different combinations: one may have high efficacy and low legitimacy, and the other may have moderate but balanced results. For management, these situations require distinct interventions. That's why the recommended reporting is not a one-time ranking, but a table where the overall index is accompanied by sub-indices, trends over time, context indicators, and data quality notes. The comparison with the previous period can be more informative than the comparison between counties, especially when the territorial differences are large and the structures have different operational profiles.

The responsible use of the model also implies avoiding the transformation of indicators into rigid targets. When staff know that the evaluation directly depends on a certain value, the activity can focus on maximizing it, even if the effect on the mission is reduced. The mechanism can be limited by combining volume and quality indicators, auditing definitions, controlled rotation of secondary indicators, and maintaining a professional assessment component. The model is designed as a tool to support decision and research,

not as an algorithm that replaces managerial responsibility. Its empirical application would require validation, piloting, documentation of sources and continuous verification of the behavioral effects produced by the measurement system.

In practice, reporting should be accompanied by attention thresholds and not by absolute labels of the perform/non-performing type. A threshold may signal the need for further analysis when a sub-index declines rapidly, when the difference from its own historical trend becomes large, or when the deterioration of several dimensions occurs simultaneously. Such a rule turns the dashboard into a warning system, not an automatic sanctioning mechanism. Management can then review documents, case distribution, and exceptional events before taking action. This sequence is important for maintaining human responsibility in the decision-making process. Indicators provide consistency and comparability, but the final interpretation must remain connected to the legal framework, the specificities of the mission and qualitative information that cannot legitimately be reduced to a numerical value.

The final interpretation should end with an explicit link between the signal, probable cause, action and the criterion by which the action will be reassessed. Without this loop, the dashboard remains a descriptive tool. Performance management becomes useful when the information leads to a verifiable measure and when the effect of that measure is subsequently observed. Thus, the model can support organizational learning and not just the periodic reporting of values.

#### 4. Discussions

The proposed framework demonstrates that the managerial evaluation of a police structure requires the separation of several levels of analysis. Context and operational pressure describe the conditions under which the structure operates, while effectiveness, responsiveness, resources, and legitimacy describe dimensions over which management can have a direct or indirect influence. Treating all these elements in a single analytical plane could produce incorrect interpretations. A structure with a high volume of requests should not be penalized just because the average response time is longer, just as a structure with a low rate of reported facts should not be automatically considered efficient.

Mathematical formalization brings transparency, as it forces the researcher to define the numerator, the denominator, the orientation of the indicator, the unit of measurement and the aggregation rules. However, mathematical precision does not eliminate conceptual ambiguity. A ratio can be calculated correctly and, at the same time, insufficiently measure the phenomenon being tracked. Therefore, each formalisation block must be accompanied by an institutional justification, a description of the data source and an analysis of the limits.

The synthetic index can facilitate the communication of a general picture, but it involves the loss of details and the introduction of normative choices. Normalisation determines the

relative position of a structure, weights reflect the importance attached to dimensions, and the aggregation method determines the extent to which high performance in one area compensates for poor performance in another. For this reason, the overall index should be published together with the dimensional sub-indices. A single value, presented without structure and without ranges of sensitivity, can create an unjustified impression of objectivity.

## 5. Conclusions

The management of police structures cannot be evaluated by a single measure or by reporting exclusively on the volume of activity. The specificity of the police mission, the diversity of responsibilities, the influence of the territorial context and the importance of legitimacy require a multidimensional approach. The literature supports the extension of the assessment beyond traditional indicators of crime and the inclusion of resources, quality of service, relationship with the public, integrity and organizational capacity.

The article proposed a framework consisting of indicators of context, effectiveness, response, human resources, material and financial resources, quality of service, legitimacy and management. The four areas of mathematical formalization show how these elements could be expressed mathematically and integrated into subscripts or a synthetic index. Formalization contributes to conceptual clarity, but does not automatically turn indicators into valid performance measures. Validity depends on the quality of the data, the adequacy of the definitions and the interpretation in relation to the context.

The main contribution of the approach is to provide a theoretical architecture that can be used for the design of further empirical research. The article does not process real indicators and does not draw conclusions about the performance of a concrete structure. The numerical examples are exclusively demonstrative, and the synthetic index is presented as a methodological option, not as a validated tool.

Future research could test the availability and comparability of data, validate indicators by consulting experts, and analyze the sensitivity of results to normalization, weighting, and aggregation. Quantitative evaluation should also be complemented by qualitative methods, such as analysis of managerial documents, interviews and examination of decision-making processes. Only the combination of these perspectives can produce a credible picture of how management contributes to the performance and legitimacy of police structures.

The use of indicators in police management can produce behavioral effects. When indicators become rigid targets, staff and management can steer their work towards maximizing the reported metrics. This risk is particularly important in the case of sanctions, controls, interventions and resolution rates. In order to limit formal optimization behaviors, the system must combine volume indicators with quality indicators, avoid mechanical reward of values and maintain the possibility of professional evaluation of the context.

The dimensions of legitimacy, integrity and procedural fairness should not be treated as secondary elements. A structure can achieve seemingly favorable operational results, but it can produce institutional costs through unfair interactions, unequal treatment, or the deterioration of public trust. From this perspective, a balanced managerial model must provide either sufficient weights for these dimensions or minimum thresholds that do not allow the full compensation of poor performance by high results in other areas.

The model remains a conceptual one. For its application, it would be necessary to precisely define the units of analysis, choose the period, validate the sources, harmonize the definitions, treat the missing values and test the robustness. It would also be necessary to consult practitioners to determine whether the selected indicators reflect actual activity and whether the formulas can be fed with available data. In the absence of these steps, the framework should only be used as a tool to structure the scientific discussion.

---

## References

---

- Hine, K.A., and Davenport-Klunder, K. (2022). From the aspirational to the tangible: Mapping key performance indicators in Australian policing.. *International journal of police science & management.*, 24(2), 382-396.
- Klose, S. (2024). e-defining Evidence-Based Policing.. *Policing: A Journal of Policy and Practice.*, 12(5), 29-38.
- Maslov, A. (2016). *Measuring the Performance of the Police.*. 2nd ed. Chicago: Carleton.
- Moore, M.H., and Braga, A.A. (2004). Police performance measurement: A normative framework.. *Criminal justice ethics*, 23(1), 3-19.
- Sherman, L.W. (2013). The rise of evidence-based policing: Targeting, testing, and tracking.. *Crime and justice*, 42(1), 377-451.
- Sparrow, M.K. (2015). Measuring performance in a modern police organization.. *Psychosociological Issues in Human Resource Management.*, 3(2), 17-52.
- Sunshine, J., and. T.T.R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law & society review.*, 37(3), 513-547.
- Tiwana, N., Bass, G., and Farrell, G. (2015). Police performance measurement: An annotated bibliography. *Crime science.*, 4(1), 26-38.
- Tyler, T.R. (2004). Enhancing police legitimacy. *The annals of the American academy of political and social science.*, 539(1), 84-99.
- Wisniewski, M., and Dickson, A. (2001). Measuring performance in Dumfries and Galloway constabulary with the balanced scorecard. *Journal of the Operational Research Society.*, 52(10), 1057-1066.