Insight in performance of safety regions

Research into the feasibility and desirability of a system of indicators for the performance of safety regions

KWINK GROEP

TU Delft

Summary

Background and remit

In 2013, the Hoekstra Committee concluded that supervision of disaster and crisis prevention could be made more effective. The Committee indicated that the Security and Justice Inspectorate, which is charged with monitoring compliance with the Safety Regions Act (Wvr), focuses too much on compliance with legal requirements and too little on the actual functioning of safety regions. The report of the Hoekstra Committee prompted the Security and Justice Inspectorate and the National Coordinator for Security and Counterterrorism (NCTV) to arrange for further research to be conducted under the guidance of the Scientific Research and Documentation Centre (WODC). The aim of this research was to gain insight into the feasibility and desirability of assessing the performance of safety regions on the basis of more output indicators or even outcome indicators, as compared to the current indicators based on parameters (compliance with legal standards, organisational requirements). The study focused on the following questions:

- To what extent and how, on the basis of theoretical and practical knowledge, might it be possible to develop a set of indicators that could capture the performance capacity of the safety regions?
- To what extent would the development of a system of indicators be desirable from the stakeholders' perspective?

Approach

The study began with a literature study on the theory relating to performance. In addition, eight exploratory meetings were held with representatives from both the Inspectorate and the NCTV, which are responsible for the management of the various safety regions. Documentary research was also carried out to chart experiences with performance measurement in the wider field of security and five other sectors (healthcare, universities, police, civil infrastructure maintenance and offshore gas). Design criteria were drawn up, based on insights from both the literature and practice, and applied in the context of the safety regions. For this purpose, meetings were held across the full breadth of the safety region domain, at the national and local levels, and at the administrative, management and operational levels. Discussions were held in three safety regions (Amsterdam-Amstelland, IJsselland and Gelderland-Zuid) and a system exercise was attended in one region (Rotterdam Rijnmond). These discussions produced a range of views on how the safety regions should be assessed and the need for and utility of performance measures. The impressions that emerged from these discussions were tested at a meeting with representatives of a number of other safety regions, confirming the views that had been expressed.

Performance measurement in the literature

A study of the literature provided several insights of relevance to the search for output and outcome indicators for the safety regions.

KWINK GROEP

TU Delft

An effective performance measurement system must include indicators that score highly in relation to the following three aspects. **Measurability, relevance** and **accountability**. The concept of 'measurability' relates to the question of whether performance can be measured unambiguously, and therefore whether it can be demonstrated that an improvement or deterioration has taken place. Relevance centres on the question of whether the performance will help to achieve the organisation's goal. In other words, does the performance measured make any difference to the organisation? Accountability relates to whether the performance measured can be attributed to a particular individual, who is formally accountable for performance in that area.

Whether the measurement of performance has beneficial or adverse effects is partly dependent on the **impact** of a performance measurement system. Is the impact high or low? A higher impact means that whether a performance measure is attained or not will have major implications in terms of holding people to account. At stake here is not only the actual impact, but also the perceived impact: in the view of the players involved, has attaining a good level of performance had a significant impact? If professionals see the system as unfair or 'stupid', there is a higher likelihood of misuse of the indicators.

If problems with measurability, relevance and accountability coincide with a significant (perceived) impact, the risk of misuse is extremely high and the incentive effect will be limited. The logical alternative would seem to be to focus performance measurement on 'learning' rather than focusing strictly on performance. However, this runs the risk of performance measurement having only a limited incentive effect. Performance measurement is most effective when a good balance is struck in terms of the strength of incentives by focusing both on 'learning' and 'accountability'.

Performance measurement in practice

We investigated experiences in the healthcare, university, police, maintenance, civil infrastructure and offshore gas sectors, as well as studying previous initiatives involving performance indicators in the safety domain. This provided insights in relation to the search for output and outcome indicators for the safety regions:

- Performance measurement and the systems associated with it are complex. There are multiple actors and functions linked to performance measurement. Holding actors to account for their performance can often be problematic.
- We did not encounter the use of significant outcome indicators in any of the sectors investigated; at best, what we saw were output indicators.
- Experience of previous initiatives involving performance measurement in the safety domain (e.g., Aristoteles and RemBrand) shows that a purely quantitative approach to performance indicators fails to achieve the desired results and creates a false sense of security.
- If the indicators in performance management systems have a large impact, the risk of perverse effects will also be significant.

- In many of the sectors studied, systems have been ‘further developed’. For example, it appears that the focus has shifted from scores based on indicators to a discussion aided by the use of indicators.

Design requirements

The insights from practice and literature are translated in design requirements. The design requirements are based on the idea that at a minimum, the set of indicators should meet these requirements in order to be meaningful and successful, and in order to prevent indicators being used in ways that produce unintended adverse effects (e.g. ‘box-ticking'). The following four design requirements were formulated:

1. The indicators in the system should be **measurable** wherever possible.
2. The indicators should be as **relevant** as possible.
3. The indicators should be **accountable** wherever possible.
4. Although measurability, relevance and accountability are important issues, the **perceived impact** of the performance measurement system for the safety regions should be limited in order to avoid unintended adverse effects.

Impressions of performance measurement from the regions

We have gathered the dominant impressions from discussions in the regions. These impressions concern how the necessity and usefulness of performance measures are seen in the safety regions.

The regions studied acknowledge the importance of information on performance, the minister's responsibility for the system and vertical accountability. Those interviewed in the regions also acknowledged the importance of performance indicators when these concern statutory requirements. However, we also encountered criticism and fear of the mechanical use of performance indicators. There is little enthusiasm for the replacement of the existing indicators by output and outcome indicators. First, this is because output and outcome performance in safety regions is difficult to express in a system of indicators. Second, it is because the low frequency and unique character of a disaster/crisis mean that the validity of indicators is limited in the hot phase (when an incident actually occurs). Third, it is because the relationship between hot and cold in a performance measurement system is problematic. Fourth, it is because there is great fear that indicators will be used for other forms of accountability, in the broadest sense of the word (for example, by means of ranking). Indicators are perceived to have a high impact.

Alternative views were also expressed, partly from the regions and partly from the Inspectorate and the NCTV. For example, the call for indicators that are better than the current process indicators also came from the region itself (Safety Advisory Board/*Veiligheidsberaad*). This creates a confused picture: on the one hand, there are calls for better indicators from the region, and on the other, no further specification was provided when requested and the regions primarily highlight the disadvantages of indicators. The region's criticism of the ranking of safety regions can also be seen from a different perspective. At the very least, rankings have

prompted policymakers to take action. The pressure to move higher in the rankings can also be interpreted in a positive light.

Conclusions

After applying the design requirements and having considered theory and practice, we conclude that it is neither possible nor desirable to apply a system of output and/or outcome indicators to assess the performance capacity of the safety regions.

- First, measurability and accountability are problematic issues. Although the examples of output and outcome indicators mentioned in the study are certainly relevant in many cases, they are seldom measurable and accountable.
- Second, safety regions have a number of characteristics that reinforce the problematic nature of the use of output and outcome indicators:
  - The hot-cold dynamic. Performance in the 'hot' phase is the most significant, but incidents are unique and rare occurrences; it is therefore not possible to assess performance in this phase using indicators. Performance in the 'cold' phase (when no incident has occurred and the focus is on issues such as training and exercises) can be measured much more easily, but it is impossible to assess performance in this phase using output and outcome indicators.
  - Mono/multi-interplay: the interaction between the separate vertical columns (mono) and the region that is meant to provide added value by linking these columns (multi). Many of the activities in the regions are inter-relational in nature (consultation, cooperation, knowledge-sharing, coordination and evaluation). The only way to assess these is by using process indicators, but these provide little information.
- Third, we encountered great resistance in the regions to the very idea of indicators or additional indicators. A range of reasons was given for this: regional diversity (which is at odds with a unified system), the relatively short history of the regions (which are still developing) and the tension between vertical and horizontal accountability (which creates unintended interference). Even without this resistance, however, it would remain the case that measurability and accountability are problem areas that would obstruct the development of a set of output and outcome indicators. There is also the risk that mistrust of these indicators could give rise to the perception that they would have an excessive impact, which could create an incentive to misuse the indicators.

The combination of problematic measurability and accountability and the perception that the indicators would have a high impact could create strong incentives for misusing the indicators. A system consisting of output and outcome indicators would have a limited incentive effect. Indeed, it would to lead to a 'box-ticking' approach and create the illusion of performance, as is the case with the current system.

Reflection

The conclusion that it would be inadvisable to develop and implement of a system of output and outcome indicators raises the question as to how the Inspectorate and the NCTV might improve oversight of the performance capacity of the safety regions. After all, it is perfectly legitimate for these institutions to seek such oversight.

We would argue that once it has been acknowledged that it is impossible to evaluate the work of the regions using a system of indicators, this might create the space needed for further productive discussions about accountability and learning. A shift is required, away from identifying new output and outcome indicators and towards 'dialogue' or 'critical interaction' between the government (the Inspectorate and the NCTV) and the regions. This would also create an opportunity to identify what really matters in the safety regions.

In order for this shift to occur, attention would need to move away from the search for output and outcome indicators and towards the design of critical interaction; that is, towards the question of (1) the appropriate substantive format for such a dialogue, (2) the rules that would apply to such a dialogue, and (3) how to deal with the tension between accountability and learning.

Some might see a shift away from indicators and towards dialogue as a transition from the hard and the challenging (indicators) towards something that is soft and easy, perhaps excessively so (dialogue). However, the opposite is true. On closer inspection, indicators are often much softer than they first appear, whereas a dialogue can assume the form of a hard challenge or a confrontation. Dialogue also involves contributions from both the Inspectorate and the region. Indicators create distance: one measures whether an indicator has been achieved and one reaches a conclusion; it could be said that no interaction at all is required. Dialogue, by contrast, involves interaction by definition. It would demand more intelligent answers from the region, and the Inspectorate would have to adopt a more intelligent approach to the issue.