Benchmarking: the missing link between evaluation and management?
by Emilio Pitarelli and Eric Monnier

Résumé
En se concentrant de plus en plus sur les résultats et en adoptant des modes de gestion inspirés des entreprises privées, le secteur public est passé d'une culture administrative à une culture de gestion, et d'une culture d'application des règles à une culture de la performance. L'utilisation du Benchmarking dans les administrations publiques s'inscrit dans cette tendance. Le Benchmarking consiste à faire des comparaisons entre un processus utilisé dans une organisation et le même processus utilisé dans une organisation plus performante, le but étant de fixer de nouveaux objectifs.


Au départ, en tant qu’équipe d’évaluation nous avions envisagé le Benchmarking comme l’un des outils d’évaluation au même titre que les autres instruments utilisés : interviews, enquête par questionnaire,….. A l’issue de la deuxième expérience de mise en œuvre de cette démarche il nous est apparu que le Benchmarking est mal adapté à l’évaluation proprement dite par contre il renforce spectaculairement les retombées de l’évaluation dans le cadre du management. En effet, grâce à la comparaison avec des "good practices", il permet de traduire des recommandations stratégiques/politiques en propositions d’actions concrètes et donne une capacité de mobilisation des gestionnaires. En conclusion, le Benchmarking peut conduire à des déceptions ou des succès selon la place qu’il lui est donné dans un processus d’évaluation.

---

1 University of Geneva and Centre for European Evaluation Expertise
Monnier@c3e.fr
Abstract

Inspired more and more by result analysis and management styles used in the private sector, the public sector has moved from an “administrative culture” to a “management culture”, and from a “culture of rule enforcement” to a “culture of performance”. The use of Benchmarking in public administrations is a sign of this new trend.

Benchmarking means comparing the processes used in one organisation with those of another considered to be gold-standard references, in order to set new objectives.

In the field of evaluation, Benchmarking is used to assess the performance of a programme by comparing it to another programme recognised as being very successful.

This paper is based on two evaluations carried out at the request of the Swiss Agency for Development and Co-operation, namely the evaluation of an environmental research and development programme, and the evaluation of the activities of the Graduate Institute for Development Studies in Geneva.

At first, we considered using Benchmarking alongside the other evaluation tools selected, i.e. interviews and questionnaire surveys. After using this new tool on two occasions, we came to the conclusion that it was not well adapted to evaluation itself, but that it substantially enhanced the impact of evaluation on management. Comparison with “good practice” allows for strategic/political recommendations to be translated into proposals for concrete action, and helps managers to mobilise their capacities. Finally, the success of Benchmarking depends on the way it is planned and used in an evaluation process.
Introduction: reform in the public sector

For the past fifteen years the public sector has been influenced by a reform trend in Europe. The sector now tries to concentrate its efforts on results and to adopt styles of management from private firms. Public services thus tend to move closer to private management, as reflected in the term "new public management". Major characteristics of “new public management” are a focus on service quality, introduction of measures and indicators of performance (for instance Benchmarking), integration of competitive principles, and efficiency in the use of resources (Cheung, 1997).

Nowadays two major issues run through the public sector: an effort to reduce public spending, and an effort to deal with the crisis in the legitimacy of administrations.

Privatisation, decentralisation, reengineering and quality are some examples of efforts undertaken to respond to these difficulties that have brought the public sector to move from an “administrative culture” to a “management culture”, and from a “culture of rule enforcement” to a “culture of performance” (OCDE, 1994).

Performance is a relative concept that can be assessed only in relation to a reference. Thus, there are different bases of possible comparisons. For example:

- What the organisation achieved in the past in relation to the present;
- What has been defined as objectives;
- What could be achieved in the current circumstances;
- What is being achieved by comparable organisations (the benchmarks).

Benchmarking

Imitating the constitutions of other countries, studying the tactics of top sports teams, spying for industrial purposes: all these activities contain the same elements as Benchmarking. Formerly Benchmarking referred rather to competitive or industrial analyses. This type of method and even the “good practices” approach could still be considered as belonging to Benchmarking, but other new elements have been added.

The broader approach of Benchmarking (Balm, 1994: 23) implies Benchmarking all key processes, and measuring oneself in relation to direct competitors as well as to non-competitors known as being the best in their own sector ("Best in the Class", BIC). Finally, it also means focusing on comparative measures that interest the service users.

---

2 All the quotations are translated from French by the authors.
"Benchmark", a term borrowed from surveyors, means a marker used as a point of reference for comparison of directions. A benchmark is therefore “a point of reference from where one can measure things” (Chang, 1995, 21).

Some managers use Benchmarking to talk about objectives setting.

“Benchmarking is a process with which a small team can help the firm to effectively embark on a continuous process of improvement and eventually to become the best. The firm reaches these goals by perfectly mastering its processes (or products and services), by finding other companies that are the best in what they are doing, by learning what the performance of those firms is with respect to criteria based on client demand, by understanding how they reach such performance levels, and by adapting the best ideas to its own processes” (Balm, 1994, 141).

The basic idea of Benchmarking is to compare a process used in an organisation with the same process used in a more efficient or effective organisation. The aim of this comparison, or calculation of variance, is to help to set new objectives.

Benchmarking is based on information exchange with firms renowned for their excellence in a given domain. It is then defined as the effort made to find and implement best practice.

Use of Benchmarking in private and public sectors cannot be understood without grasping that it is fully part of a quality management approach, in particular the Total Quality Management (TQM) approach. Benchmarking aims for continuous improvement of processes and client satisfaction, even though it is an indirect approach because it consists of observing what other service providers do. However, Benchmarking is only one of many ways to improve public service quality.

“Any public service organisation has to ask the following question: can there be close correspondence between specific features of Benchmarking as a technique and the characteristics of the service to benchmark?” (Pollitt and al., 1994, 20).

Initially developed in the United States in the 1970s, TQM was used extensively by the Japanese before being introduced in Europe. With the adoption of international ISO standards the concept was officially accepted on our continent.

Quality cannot be obtained without measures. Self-evaluation is a way to systematically examine an organisation's activities and results, compared to models of excellence which are used as benchmarks. The Deeming Prize created in Japan in 1951, the Malcolm Baldrige Prize founded in the United States in 1987, and the European Quality Prize launched in 1992 are the most famous models of quality.
These prizes are beginning to be won by public services: hospitals, municipalities, etc. The main elements taken into account are policy and strategy, staff involvement and client satisfaction.

Quality is a philosophy that has entered into the world of private and public management, of which communication, training, Benchmarking, performance measurement and certifications are the main tools.

**Quality is an objective, quality management (TQM) is the means to attain it, and Benchmarking is one of the tools in this global approach.** In modern business administration, analysis is based on process operation, in other words, on the **way things are done**.

Three types of process exist. *Key processes* participate directly in the realisation of the product or service: they are transversal. An example is a customer service. *Support processes* are necessary in the realisation of key processes, even though they are not directly apparent for users. Data processing and bookkeeping are typically support processes. Finally, the main object of *management processes* is to steer the organisation.

Thus, **Benchmarking consists first in breaking up a process into activities, then in measuring those activities, and finally in comparing them.**

The most important types of Benchmarking are as follows:

*Internal Benchmarking* is the practice inside an organisation itself. In large firms, similar processes exist in different departments, so that internal comparisons can be made. For example, a large firm can compare processes such as customer services or repair services in different branches of activity. Potential partners are normally easy to identify, but there are few opportunities to find important sources of improvement.

*Competitive Benchmarking* refers to an organisation that excels in a particular process. Identical or similar activities are compared. Two cantons\(^3\) can, for instance, compare their processes in the social service domain or two communes their road maintenance practices. Benefits of this approach are potentially important. One still has to be sure that there is no problem of sharing confidential information, and that data can be compared.

---

\(^3\) Switzerland is divided into 26 cantons which are subdivided into communes.
Finally, some functions or processes are identical irrespective of the differences between sectors. *Generic Benchmarking* consists in comparison with the best firms that have similar processes, irrespective of the field of activity. For instance, a department of social affairs could study the reception area of a bank. It is more difficult to identify such organisations, but the improvement potential is considerable.

In general, different types of Benchmarking are combined (mixed or hybrid Benchmarking), and the comparison involves several firms.

In light of this presentation, it appears that **there is no theoretical formulation of this technique**. Outlines are vague, and the minimal elements allowing one to speak of “Benchmarking” seem to be a comparison from process to process, a contract (because there is real sharing of information), and an objective of improvement through best practice.

Implementation of Benchmarking takes place in **stages** where each stage has its importance. It is therefore advisable not to begin the comparative analysis itself too early, before the following preparatory steps have been taken: determining the products of importance to the organisation (what), choosing whom to compare with (who), and deciding on data collection (how).

The Benchmarking approach consists primarily of:

1. Selecting what has to be benchmarked (what has to be improved);
2. Identifying Benchmarking partners (benchmarks or points of reference);
3. Collecting and analysing internal and external data;
4. Determining competitive differences by comparing internal and external data;
5. Fixing future feasible performance levels (new objectives);
6. Communicating Benchmarking results;
7. Elaborating action plans;
8. Setting up concrete actions (project management).

Conditions for successful Benchmarking are similar to those required by most improvement methods, such as TQM: involvement of management, planning and project management, use of tools facilitating team work, adequate training policy.

**Benchmarking and public management**

The public sector has been under pressure to improve performance for several years. Various methods aimed at such improvement exist, including Benchmarking.
In the public service, Benchmarking first of all allows staff to know their results in relation to others. This can motivate them to improve service quality. Colleagues, faced with their own processes, are thus integrated into a learning process (Baumann and Trepp, 1996). We can consider, for instance, medical and educational professions (e.g. physicians, nurses or teachers) which rely on strong national and international networks where ways of doing things are open to discussion.

Another motivation to apply Benchmarking lies in the improvement of information on public services. Comparison allows the definition of a sort of norm which will be used to publicly display particular aspects of the service supply.

Organisations providing public services are organised in strata in Switzerland (Confederation, cantons, and communes). Benchmarking can therefore be used by a higher level to fix objectives for a subordinate entity.

One of the strongest incentives to use Benchmarking in the public sector is still the possibility of introducing competition. Managers can use it to indicate the level of performance reached, or to look for best ways to supply a service. Benchmarking also allows an opening towards the outside: the organisation has to admit that it is not unique, and that there are sources of inspiration to improve its performance (change factor).

Benchmarking is a technique that can be used to formulate answers to organisational problems and to highlight strong and weak points. One of the difficulties stems from the absence of comparative information on the performance level of other public service organisations, and it is sometimes useful to call on expert panels to better identify best practices. Moreover, processes are not always clearly specified, and clients are difficult to identify.

Behind the internal processes of an organisation there is a series of factors (e.g. staff, offices, technology, budget). If these cannot be controlled, then the comparison basis will be distorted (Pollitt and al., 1994: 15). In international Benchmarking, the problem of weighting heterogeneous factors is increased. Cultural values will also influence users’ expectations and therefore their requirements.

The success of Benchmarking depends on the quality of the selected performance indicators. Without performance indicators, Benchmarking remains a relatively unstable exercise. However, these indicators can distort behaviours: teachers teach their students to pass exams instead of educating them, state employees maximise their results on the performance indicators chosen instead of providing a service.
A way to reduce the capacity that performance norms have to distort behaviours is to link indicators to what is important for end users. Australian post offices, for example, chose as an indicator the number of letters sorted in a day. As a result, pre-sorted mail was sorted again so as to improve apparent performances. In the end, that indicator was replaced by the proportion of mail delivered in time from mailbox to addressee.

In Switzerland, Benchmarking serves primarily as an evaluation tool for performance in management of the communes. This can be explained by the fact that communes generally supply the same services and that they have a relatively reduced size, which facilitates comparisons.

The advantage of comparison lies in the fact that most communes encounter comparable problems. The setting-up of comparative elements to position communal administrations according to their main activities can be used to highlight possible needs for action, which results in implementation of improvement measures. Benchmarking, by detecting the weak points of communes, is used as a starting point for future reforms.

Benchmarking seems to suit commercial public firms particularly well, such as electricity or water suppliers, purification stations and incineration plants. Most of these firms effectively observe market economy rules.

Other administrative branches also lend themselves to comparative analysis.

The Corseaux Beach swimming pool example speaks for itself (CES, 1998, 3). Comparison between swimming pools of different communes allowed the Vevey commune to confirm the existence of considerable potential to save energy at its swimming pool. Without reducing users' comfort, arrangements were made (switching off the ventilation during the night, reducing the working time of pumps, etc.) to improve the mode of operation. Substantial energy savings were thus achieved.

Another example concerns the comparison, aided by indicators, of the cost per inhabitant of different services in selected communes of the Argovie canton in Switzerland (Güller, 1998). By analysing results, it turns out, for instance, that Aarau devotes twice as much per inhabitant as Baden to social assistance. The question is then whether that difference is justified. Moreover, sound knowledge of the context is important before judging the situation. The Brugg swimming pool, for example, spends twice as much per inhabitant as Aarau. Before deciding which measures to adopt, one has to know the situation: entrance cost, number of inhabitants in the region, opening time, free staff, etc.
Finally, the stakeholders (staff, public authorities, and users) do not have the same quality objectives, so that different performance criteria are required.

Difficulties of Benchmarking transposed onto the public sector lies in the fact that a quality culture does not exist in public administrations. Who is the best in the public sector? Most of the time, it is necessary to carry out a fairly extensive inquiry to answer that question.

Benchmarking applied to public administrations should be considered as an element of a new performance culture. Most countries take the private sector as a reference, as it seemingly offers more stimulating standards and higher credibility among opinion. Thus, experiences of real Benchmarking in the public sector still remain limited.

The necessity for doing international comparative analysis is favoured by several current factors:
- Globalisation and increasing dependency (e.g. financial questions, environment, communications);
- Lack of national comparison points;
- Wish to broaden debate and thought.

**Benchmarking and evaluation**

Benchmarking is often seen as one of the new evaluation tools. It consists of finding measures recognised as having functioned successfully (best practices) and using them as comparison standards. This serves to clarify the programme logic and to identify strong and weak points. Judgement criteria selected are above all effectiveness and efficiency. The advantage of Benchmarking is its ability to detect good practices and to adapt them so that they can be applied in other situations.

Benchmarking concerns two types of evaluation. *Operational evaluation* concentrates on questions of implementation. Characteristics of the environment and its evolution are taken into account. *Impact evaluation* examines the real effects of a policy and focuses primarily on the achievement of objectives.

Contributions of *comparative policy analysis* (Knoepfel, 1992) are important here. A policy generally follows different stages: policy conception, administrative programme, implementation, control and evaluation. Policy analysis teaches us that the failure of a policy can result not only from the programme itself but also from deficient implementation. As a
result, implementation can be defined as an important element between planning (design) and effectiveness (effects).

Research on implementation therefore offers access to research on effects. It aims to underline regularities by delimiting the operation of a policy. Various methods are applicable in this case: documentary analysis, interviews and Benchmarking. In practice, however, it appears difficult to extend this research to the study of effects.

In these conditions Benchmarking, in other words the comparative analysis of several implementations, appears as a very appropriate tool to study the development of a programme.

While evaluation consists above all in assessing policy effects, it will often demand sound knowledge of the processes that have produced the observed effects. This implies the need for an analysis of the policy design and implementation.

“It is not sufficient to establish an exhaustive inventory of effects; one still has to explain how and why they happened, if they are the consequence of legislation itself or if they reveal modes of implementation. The distinction is important, for it allows observed results – effective or not – to be attributed to qualities or defects of legislation or its modes of implementation.” (Horber-Papazian, 1992, 151).

In practice, evaluation cannot be blind to the internal mechanisms of a policy. This leads us to say that Benchmarking, even though it does not evaluate effects, enables us to analyse the internal processes of a programme.

Benchmarking appears as a particularly suitable tool for operational evaluations. This amounts to comparing various ways of implementing programmes (or projects) in order to improve the implementation of the evaluated programme. This way of proceeding makes the quality improvement of the provided services possible.

Benchmarking can theoretically also serve impact evaluation. In some cases, Benchmarking is used to compare the effects obtained by different organisations or policies. However, possibilities for obtaining reliable results are limited, because effects are generally tangled and because factors having an influence on results, for example implementation, are numerous.

Benchmarking poses a challenge: how to find relevant criteria for evaluating and improving activities?
Case study

Our analysis is based on two evaluations entrusted by the Swiss Agency for Development and Co-operation (SDC) to the Centre for European Evaluation Expertise (C3E) in Lyons: evaluation of the Swiss environmental priority programme, and evaluation of the Graduate Institute for Development Studies (IUED) located in Geneva. For these two evaluations, an approach close to Benchmarking was applied. We wish to show how Benchmarking can contribute to enriching policy evaluation.

The first case concerns the “development and environment” module of the Swiss Priority Programme – Environment (SPP-E). Following the Rio Summit, environmental problems such as population growth or climate changes have been considered as global. The “development and environment” module (Module 7) was initiated to enhance southern countries' abilities in environmental research. To achieve this goal, it created research collaboration between North and South and plans for financial and technical assistance.

As the programme approached an end, the Swiss authorities decided to launch an evaluation in April 1999. This would help to plan future activities in Switzerland in the field of North-South research collaboration.

The objective of this ex post evaluation was to draw lessons from many years of practice and particularly to test new research collaboration approaches, in order to guide future decisions. The evaluation included effect assessment in relation to set objectives and to international benchmarks.

Evaluative questions were the following:

- Did the programme strategy and concept make sense with regard to international needs for research capacity in the South?
- Did the programme mechanisms achieve an optimal balance between capacity building, scientific results, and cost?
- In which cases did project management produce the best outcomes (in terms of both research capacity and scientific results), and how did it work?

The evaluation included a comparison of Module 7 with international good practices in the field of North-South partnership research.

The Swiss authorities wanted to compare the programmes in the following respects:

- Monitoring and management system;
- Communication between programme managers and researchers;
- Construction and management of North-South collaboration;
- Project selection;
- Collaborations sustainability.

This exercise involved three experts from national or international agencies with extensive experience of North-South research: a professor from the national University of Benin, an official from the evaluation unit of the International Development Research Centre in Canada, and a member of the Dutch advisory committee on scientific research in developing countries.

In July 1999, these experts received the intermediate evaluation report. They examined all data and analyses in their final state, collected by questionnaire and interviews, plus three case studies. We then asked them to compare the Module 7 management and results, as they appeared in the evaluation material, with management and results of other North-South cooperation “success stories” which they considered to be comparable. The experts submitted their written comments before the steering group discussed the conclusions and recommendations of the evaluation. Two experts participated in the final meeting held in Bern in August 1999.

International comparisons produced interesting results, particularly in project selection, partnership sustainability and management. They also played an essential role in the formulation of recommendations. On the other hand, the exercise was not very conclusive concerning the judgement function (of evaluation as such) that we expected to get out of it. Learning from this first attempt, we considered Benchmarking not as an evaluation tool as such but as a complementary exercise at the end of the process.

The second case concerns the Graduate Institute of Development Studies (IUED). An agreement with the University of Geneva formalises relations between the two institutions, while guaranteeing the autonomy of the Institute which has been managed since 1975 by a private foundation. Its resources are made up of State of Geneva and Swiss Confederation subsidies (mainly SDC), and by fees for programme management or execution of mandates (e.g. missions, expertise, training).

Since 1992, the Institute has been engaged in a restructuring and programme diversification process in order to adapt to new challenges, for the international context has evolved considerably: environmental risks, market deregulation, increase of humanitarian crises. The division between North and South is not as clear as in the past, and the opening of Eastern European countries has generated new issues in terms of development. All these changes influence thinking about development: teaching, research and operational activities all have to take them into account.
Moreover, Swiss academic institutions are confronted more and more with budgetary pressures and the necessity to better integrate themselves into the international academic community.

Evaluation was applied to IUED activities. Since its creation, the IUED's activities had been evaluated twice: in 1976 and at the end of the '80s. The current evaluation concentrated on the last ten years, and intended to formulate useful conclusions for years to come.

In order to specify evaluative questions, the whole steering group performed an exercise aimed at clarifying the IUED intervention logic. This was used to draw a diagram of expected impacts showing relations of cause and effect.

Following this teamwork, the steering group selected four evaluative questions:

- How does the integration of research activities, teaching and operational project implementation represent a comparative advantage for the IUED in relation to other institutions?
- How do the fame, activities and partnerships of the IUED contribute to Geneva's and Switzerland's influence?
- To what extent have IUED activities and results made a significant contribution to strategic reflection of national and international decision makers, notably to devise, promote and launch development policies adapted to new and evolutionary requirements of the development sector?
- To what extent do the results of IUED activities and services meet sponsors' expectations?

It was decided to complement the evaluation system by a Benchmarking exercise in order to improve IUED performance in certain areas. Domains selected for comparison were:

- Synergy between the different types of expected results (teaching, research, and operational implementation of projects);
- Evolution of strategic guidelines;
- Impacts in terms of fame.

Each subject of comparison was examined by a chosen person in an institution considered to be the most relevant because of its performance in a domain. Three institutions were chosen to serve as standards: one in England, one in Canada and one in Holland. A file for each domain was constituted and sent to the reference institution. The exercise was based on the first evaluation results, drawn from questionnaires, interviews and focus groups. The selected peer (for instance the head of communication, responsible for fame) submitted his comments by comparing his institution with what was done at the IUED. Three managers participated in the final evaluation meeting of the steering committee and presented their comments.
In this case, the approach was more fruitful, for people were asked to show how their own institution was doing better than the IUED in some domains. It was not about judging the IUED, but showing how to do better. We were then clearly in a Benchmarking approach, and not in an expert panel as in the first case.

**Conclusion: lessons learned**

At first, we considered using Benchmarking alongside the other evaluation tools selected, i.e. interviews and questionnaire surveys. After using this tool on two occasions, we came to the conclusion that it was not well adapted to evaluation as such since the tool is unsuitable for comparing outcomes. Nevertheless, Benchmarking considerably enhanced the impact of evaluation on management. Comparison with “good practice” allows for strategic/political recommendations to be translated into proposals for concrete action, and helps managers to mobilise their capacities.

**Benchmarking is relevant mainly at the stage of formulation of recommendations** to improve implementation (the search for good ideas). Through comparisons, this tool has the faculty to underline good practices, thus enabling the people who participate in it to find new ideas for improvement. Moreover, it stimulates self-reflection by administrative authorities by specifying processes and performances. Benchmarking thus participates in the learning process that evaluation tries to produce.

The success of Benchmarking depends on the way it is used in an evaluation process. The choice of domains is crucial if the evaluation approach is to be relevant. Evaluators must define with the stakeholders the domains that will be reasonably relevant in relation to evaluation recommendations. It is advisable to concentrate on the key processes in the programme or organisation so that Benchmarking is really of use in solving problems and as an evaluation tool. This has been the case of the IUED evaluation, where management domains have corresponded to evaluation domains (strategy, communication, and synergy).

Benchmarking has thus been stimulating for the IUED, which has carried out in-depth reflection on the three selected processes. An important contribution of this tool applied to evaluations remains its capacity to indicate potential process development to people who perhaps never paid attention to it, and its role in helping organisations to consider their own practice more objectively by finding out what is done elsewhere.
One still has to remember that Benchmarking is an exercise demanding an important investment in time (between six months and one year) and in money (one person employed full time to implement it, with travelling expenses, interviews, etc.).

Since evaluation is not, strictly speaking, a scientific discipline, it can be learned only through an accumulation of experience. It is therefore important that lessons be drawn from progress made and from difficulties encountered by evaluators. Both cases presented participate in this learning process.

Bibliography

Books:

• Swiss Agency for Development and Cooperation (1999), *Evaluation of a research capacity building project*, Bern.

Web sites:

• http://www.oecd.org/puma (site presenting Benchmarking applied to public management)
• http://www.zed.ch (site of the Centre for Public Service Evaluation, which uses Benchmarking in the framework of commune management in Switzerland)
• http://www.hmce.gov.uk/general/about/bmbackg.htm (example of public sector in Great Britain, which has introduced Benchmarking as a tool for improving quality management).