Benchmarking and performance measurement in public sectors
Towards learning for agency effectiveness

Alexander Kouzmin
University of Western Sydney, Nepean, Australia
Elke Löffler and Helmut Klages
Post-Graduate School of Administrative Sciences, Speyer, Germany, and
Nada Korac-Kakabadse
Cranfield School of Management, Cranfield, UK

Keywords Benchmarking, Information technology, Learning, Performance assessment, Quality assurance, Self-assessment

Abstract Given the prevailing emphasis on agency performance, customer focus, stakeholder's interests and other methods of assessment under new public administration and prevailing managerialism in many public sectors around the world, administrative practitioners have taken to benchmarking as an instrument for assessing organizational performance and for facilitating management transfer and learning from other benchmarked organizations. The introduction of benchmarking into the public sector is still in its early stages. Technical problems, scepticism about usefulness and the appropriateness of transferring putative private sector competencies into public administration and the resistance in accepting organizational change as a necessary consequence of benchmarking exercises in the public sector, prevent the widespread acceptance and use of benchmarking in public sectors, arguably “punch-drunk” with systemic change. Nevertheless, there are some encouraging examples of benchmarking within the public sector. This paper critically analyzes these examples in order to establish the vulnerability points of such measurement instruments which, possibly, need more research in order to establish the specific learning dimensions to benchmarking and to illustrate the importance of such benchmarking and learning within the highly risky, information technology (IT)-driven experiences of systems development and failure.

Trends in performance measurement in the public sector in the 1990s
Performance is a key word permeating all discussion about “new public management” (OECD, 1993, p. 7). Part of its attraction is that performance is a broad concept: it has various meanings, for different audiences, in different contexts (Carter, 1991). This makes the design of performance indicators (PIs) in both the private and the public sectors very difficult. Besides the technical problem of operationalizing an abstract concept, the same set of PIs may need to be used to answer questions about the different dimensions of performance.

Whereas performance measurement in the private sector is, in general, seen as something normal – the assumption being that the private sector is imbued
with a performance-based culture – conventional wisdom suggests that there are special characteristics of the public sector which make performance measurement inappropriate or, at least, very difficult. Two explanations are commonly used to explain the differences in public/private performance measurement (Carter, 1991). The first assumes that because private firms putatively adhere to bottom-line profit requirements, performance measurement is a straightforward and contestable technical procedure.

The second argument focuses on the particular social and political pressures on public sector agencies. Public services operate with a fixed budget and consumer groups are in competition with each other for scarce resources. The market solution to this situation is to introduce the user pays principle in selected public services so that users of a service are the actual people paying for that service. But the imperative of the welfare state precludes this market option from economic textbooks. This problem of scarce resources implies for performance measurement in the public sector that a certain degree of insensitivity to consumer demands is positively desirable in order to protect the interests of those vulnerable consumers, least satisfied with services delivered and with the least resources for either “exit” or “voice” modes of protest (Klein, 1984). In other words, consumer satisfaction cannot be the only, or dominating, dimension in performance measurement in the public sector and has to be handled with considerable caution (Swiss, 1992).

There are also different dimensions to performance measurement. Whereas in the 1980s the focus was on the “three Es”, economy, efficiency and effectiveness, in the 1990s attention has shifted to quality and consumer satisfaction. One reason for this can be seen in a value change to the phenomenon of the “difficult citizen” (Klages, 1994); one who wants to be the “subject of his actions”. This value change involves the “difficult citizen” not passively accepting the state *per se*, in the sense of the “older” theories about the functions of the corporatist state, but taking a consumerist and “instrumental view” (Klages, 1993) of the state in the sense of public services delivered.

Three major trends in performance measurement, in OECD countries in the 1990s at least, can be identified (Australia, MAB-MIAC, 1993):

1. the development of measurement systems which enable comparisons of similar activities across a number of areas (benchmarking instruments, such as citizens charters and quality awards);
2. efforts at measuring customer satisfaction (citizen surveys; output as indicators, such as the number of complaints; and throughput measures such as indirect proxies for measuring direct impact of programs on clients); and
3. some lessening in the focus on the long-term impact of programs, particularly in evaluating such programs.
While these trends help to establish a performance-based culture in the public sector there are problems related to the lessening of attention to the long term and with the dominance of the consumerist orientation in performance measurement. In the “new public management” paradigm, the basic question which drives performance measurement in the public sector is not “how much” government one has, but “what kind” of government (Osborne and Gaebler, 1992). By concentrating performance and evaluation exercises on the means instead of the ends of public agencies, one risks measuring “peanuts” in public management.

**Major public sector performance measurement concepts**

**Benchmarking**

Benchmarking can be seen as an important management tool of total quality management (TQM). It was first developed by Xerox Corporation in 1979, when severe quality and costs problems became visible in the face of the extremely low price of Canon copier machines (Horvath and Herter, 1992). Today, this instrument is used by a large number of US companies, such as Motorola, Ford, GTE, IBM, AT&T, Honeywell and Alcoa.

Benchmarking is a term which was originally used by land surveyors to compare elevations. Today, however, benchmarking has a narrower meaning in the management lexicon since the benchmark is industry best-practice and is not in any sense a standard. Camp (1989, p. 10) defines benchmarking as “the continuous process of measuring products, services and practices against the toughest competitors or those companies recognized as industry leaders, (that is) ... the search for industry best practices that will lead to superior performance.”

The aim is to identify competitive targets which render the weak points of the benchmarking organization visible and to establish means of improvement. In other words, the basic idea behind benchmarking is not to find out “by how much others are doing better but, rather, how they make it to do better in certain areas” (Horvath and Herter, 1992, p. 5).

Given the fact that industry best practice for a given product, service or process may never be found, because of high transactions costs, only relative or local optimums are found as benchmarks. In the real world “ideal-type” definitions of benchmarking need, however, to be modified. Consequently, benchmarking is a continuous, systematic process of measuring products, services and practices against organizations regarded to be superior with the aim of rectifying any performance “gaps”.

With regard to the organization being compared, the object and the targets to be improved, different forms of benchmarking can be distinguished (see Table I).

Benchmarking is a more comprehensive exercise than “reverse” product engineering which only focuses on the analysis of specific components and functions of the products of competitors (von der Oelsnitz, 1994). Nevertheless,
in industry, reverse product engineering is often the starting point of a company’s benchmarking exercise. Companies often concentrate on the comparison of costs, as was the case with Xerox (Tucker et al., 1987). The extension from the comparison of products to the analysis of methods and processes implies that comparisons can be made with companies of other sectors that “excel” in their methods or processes. With the focus on methods and processes being compared, new targets such as time, quality and customer satisfaction emerge. The practical advantage of comparing with non-competitors is that information can be obtained much easier since competing organizations naturally have a common reluctance in sharing commercially sensitive information. Benchmarking against competitors may also uncover practices that are unworthy of emulation. While competitive benchmarking may help a company to unravel the competitor’s performance, it is unlikely to reveal the required practices needed to surpass that competitor’s performance (Camp, 1989).

The fact that most companies develop their benchmarking analysis in subsequent stages shows the complexity of this “strategic-competition/analytical planning instrument” (von der Oelsnitz, 1994). What is important for the introduction of benchmarking is that this instrument needs to be permanently used by staff (Walleck et al., 1991).

One major issue is the need to decide on appropriate indicators to be used in the benchmarking process. Financial indicators rarely exist within public agencies. Furthermore, the search for the “best of the class” is the most difficult part of the benchmarking exercise. A systematic search involves high cost and since, in most cases, only secondary information sources are available, comprehensive searches do not necessarily lead to comparable outcomes.

Another problem is to get the data needed for the analytical part of the benchmarking process. This is easier when comparisons are made “across-the-border” since companies more readily release information they would not, normally, give to direct competitors. The success of benchmarking depends on employees understanding the results (and the consequences) of the benchmarking exercise and that they will need to participate in determining and implementing necessary organizational change. Possible new performance targets have to be set and actions plans made. In markets that are changing

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Products</th>
<th>Methods</th>
<th>Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Object of benchmarking</strong></td>
<td>Products</td>
<td>Methods</td>
<td>Processes</td>
</tr>
<tr>
<td><strong>Target of benchmarking</strong></td>
<td>Costs</td>
<td>Quality</td>
<td>Customer satisfaction</td>
</tr>
<tr>
<td><strong>Reference of comparisons</strong></td>
<td>Intra-departmental competition</td>
<td>Constituencies and clients</td>
<td>Same agency or sub-unit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table I.** Forms of benchmarking

*Source:* Horvath and Herter (1992, p. 7)
fast, today’s “best of the class” may not be tomorrow’s as one has subsequently learned from the In Search of Excellence (Peters and Waterman, 1982) example. Benchmarking, therefore, has to be a permanent exercise, even tracking corporate and agency failures, in order to help establish objects, indicators and target companies to be benchmarked. Needless to say, such indicators and selected companies and agencies need to be constantly validated.

The search for the “best of the class”, the definition of “good” indicators and data collection turn out to be critical elements of the benchmarking process. It is striking that only large companies can afford to develop their own benchmarking. Other companies rely on previously developed programs (for example by participating in quality awards) or, at least, investing in external counselling (by participating in benchmarking training workshops).

Benchmarking in the public sector is analogous to that in the private sector but the motivational forces and obstacles are somewhat different. Benchmarking introduces cooperation in the private sector that, subsequently, motivates competition for market share. In the public sector, according to the paradigm of “new public management”, benchmarking is supposed to introduce competition into a state apparatus context that is characterized by the cooperation of public sector agencies for the “collective” public good. Meaningful competition is only possible between producers of the same products and services. If benchmarking is supposed to introduce competition into the public sector it has to be done, it is argued, between public agencies with very similar goals and other organizational characteristics so that actors actually perceive differences or qualitative improvements in delivering similar services to constituencies. A health agency, therefore, finds it difficult to be compared with a local municipality so that “across-the-border” benchmarking in the public sector is not as useful an exercise from the point of view of competition and its putative benefits on actor performance (Dixon and Kouzmin, 1994; Dixon et al., 1996).

There are also methodological considerations against “across-the-border” benchmarking in the public sector. Whereas a state health agency may learn innovative practices from a local municipality, it may not have the degrees of freedom to implement consequential structural or procedural improvements within its own sphere of jurisdiction. Functional benchmarking leads to frustration because information obtained from non-related public sector organizations cannot be used. This may be the explanation for the empirical finding of a survey of the Western Australian public sector that public organizations tend to compare related government agencies more than any other sources (Frost and Pringle, 1993). As far as the availability of information is concerned, there is no need for “across-the-border” benchmarking since public agencies have no competitive drawbacks to fear from passing information on to peer organizations.

It is difficult for public sector organizations to identify the “best of the class” experience themselves. Since simple financial indicators such as revenue per employee and inventory returns are missing, it is very difficult to establish
which organization can be labelled “successful” and be used as a benchmark. With respect to this problem, Quality Awards and Citizens’ Charters fulfil a very important function by delivering a benchmark for public sector organizations. The benchmarking instruments serve to identify superior performance and to make their superior practices visible to other organizations. Quality standards such as DIN ISO 9000-9004 can be interpreted as a precondition to benchmarking because it helps private and public organizations alike to define their own quality assurance system. DIN ISO 9000-9004, citizens’ charters and quality awards are measurement instruments setting benchmarks for different performance levels. Figure 1 visualises the context of the instrument of the benchmark family.

The quality assurance system DIN ISO 9000-9004

The DIN ISO 9000-9004 is an internationally recognized benchmark for quality management. It gives indications to companies as to how to develop quality management and a quality assurance system and gives a standard to external and internal audits to assess the degree of quality management of companies. The implementation of DIN ISO 9000-9004 in companies is an important step to TQM and such standardization is not as contradictory to a TQM approach as the literature often suggests (Antilla, 1993).

Underlying the mysterious formula, DIN ISO 9000-9004, is a synopsis of norms which needs to be understood as a guideline, as a benchmark for companies to improve their individual quality management but also being suitable to their needs and organizational characteristics. It is by no means a ready, of-the-shelf solution or a regulatory system in a negative sense.

With regard to the content of DIN ISO 9000-9004, it can be divided into three sections (Bläsing, 1992, p. 27):

1. instruction to use, selection criteria (9000);
2. guidelines for the development of quality management and one’s own responsibility (9004); and

![Figure 1. Benchmark instruments and performance level targets](source: Löffler (1996, p. 147))
proposals for the selection of elements in the framework of contract-based negotiations (9001, 9002, 9003).

It is also apparent that the ISO-system differentiates between the contracting and non-contracting situation. The non-contract situation is the starting point for the development of TQM for which DIN ISO 9000-9004 delivers the mode of operation. A quality assurance system consists of 23 elements, out of which a selection is proposed for the formulation of contracts. For trading in the internal European Union market, for example, this is an important marketing advantage. This also explains the increasing importance of certification for trading companies (Meier, 1993).

Does DIN ISO 9000-9004 also make sense for public sector organizations? To date, the application of the ISO-system to the public sector is still at an experimental stage. The Finnish Association of Local Authorities has done pioneering work by developing, together with five municipalities, an ISO Handbook for municipal authorities. Such a handbook, written in a language that fits the administrative culture in Europe, could be a starting point providing a minimum standard of quality in public services in the European Union. Nevertheless, the ISO-system focuses very strongly on quality and omits all other dimensions of performance measurement. For industry, in a buyer’s market, quality is an important competitive advantage; in government, however, the incentive structure is such that political success is more important than quality-oriented management (Sensenbrenner, 1991). Quality in the public sector is always desirable, in the sense of competence in the routine activities of the organization. It may not always be desirable for various stakeholders in the sense of quality being considered as a luxury. In poor constituencies, for example, taxpayers may prefer to have a poorer quality of drinking water and, consequently, pay lower water rates.

Citizens’ charters

Charters are official frameworks for assessing and awarding quality in the public sector. Whereas DIN ISO 9000-9004 measures meso-quality as an “external quality concept which applies to producer/consumer relationships” (Bouckaert, 1992, pp. 4-5), charters address macro-quality as a “generic system concept which applies to the public service/citizen relationship” (Bouckaert, 1992, p. 5). The essential idea behind charters is to increase the quality of life in society and to pay more attention to the needs of citizens. The ultimate purpose is to renew citizen trust not only in public services but also in the State. The method is to renew the Rousseau-like “social contract” (Broome, 1963) between the rulers and the ruled, by developing charters. Belgium, France, the UK and Portugal have now considerable experience with designing and developing distinctive and operational charters.

Charters sound very convincing at first sight but there are two in-built contradictions in the basic concept. The first inconsistency of citizens’ charter
is that it makes allusion to the English Magna Charta Liberatatum of 1215, which was a letter of privileges for noblemen rather than a catalogue of liberal rights for all citizens (Löw, 1986). Nevertheless, it was an important milestone in the history of human and basic rights. At the centre of this early codification of liberal rights is the citizen; a citizen being defined as a concentration of rights and duties within a constitutional state, within the rule of law, and a hierarchy of laws and regulations (Bouckaert, 1995). The “client” is a much more limited concept since the citizen is part of the social contract, whereas the client is part of the market contract. When charters are used in a new ideological context, such as with “new public management” (Kouzmin et al., 1997), the meaning of charters changes significantly. It is no longer a catalogue of rights and duties of the ruler and the ruled, but a “quality” checklist for clients of public services. This is especially true in the British Citizens’ Charter which builds on the user-service provider relationship instead of citizen-state relationships (Bouckaert, 1995).

Second, charters are based very much on older social contract theories (Rousseau, Hobbes, Locke) according to which citizens believe, unconditionally, in the legitimacy of the state as long as the state provides security (Hobbes, 1966). This is no longer true today where the market contract becomes more dominant than the classical social contract.

The Belgian French, Portuguese and British Charters are clearly influenced by contextual macro-variables such as administrative culture, the political-economic system and corporate interests (Bouckaert, 1995). The British Citizen Charter basically adds to the 3Es of economy, efficiency and effectiveness, a quality dimension through competition (UK Prime Minister, 1991). The French and Belgian reasoning, however, bases very much on the idea of accountability of civil servants to the citizen. The Portuguese Public Service Quality Charter also focuses on the idea of accountability of the state to the citizen but also on value for money. Thus, even though the charter concept is basically the same, the operationalization of citizens’ charters is very much different in the various OECD countries. Citizens’ charters are consistent with their public sector environments and are not just an ideological overlay on developed and mature administrative systems.

There is one particularity of the British Citizen Charter which is the Charter Mark Award. Applicants – all kinds of public sector organizational units – have to prove to the Prime Minister’s Citizen’s Charter Advisory Panel that they meet the Citizen’s Charter principles for delivering quality in public services (UK Prime Minister and the Chancellor of the Duchy of Lancaster, 1992). The Charter Mark winners then use the Charter Mark on their products to show that their achievements have been recognized. The Charter Mark Award is a special kind of quality award lying at the minimum end of the spectrum between DIN ISO 9000-9004 and quality awards, emphasizing optimal levels of agency performance.
Quality awards
Before analysing quality award competitions, one has to make clear what this label means; there is a whole variety of quality award competitions in the private as well as in the public sector. A public quality competition award is defined as a performance measurement instrument which fosters innovation and quality in the public sector by the identification of excellent public organizations by independent panels and with active participation of public agencies. These processes render the success factors of excellent administrative practices transparent. This definition excludes awards that are given to organizations for outstanding past achievements without using performance indicators and without submitting to a competitive selection process. For example, the grant of the Distinguished Service Cross to public organizations in Germany does not fall into the quality award category because the selection is not by a competitive process. Nevertheless, the above definition is broad enough to include innovation awards in the public sector. As Borins (1995) points out, the difference between specific and general quality awards is gradual rather than fundamental. Given the elusive nature of innovation in the public sector, judges of specific innovation awards will need to include effectiveness in overall judgements.

Quality award competitions come from the private sector and have been transferred to the public sector in the process of the paradigm shift taking place in “new public management” of many Anglo-American and European economies. The basic assumption, again, involves the putative benefits of competition: why not introduce surrogates of competition into public sectors which lack the “discipline” of market competition? The competition among the participants of an award program is supposed to motivate other actors of public agencies. However, due to specific characteristics of the public sector, quality competition awards have another important function. Whereas in the case of private goods there is a price, which is a complex measure of all the quality characteristics economic actors attribute to such goods, there is no such indicator in the case of public goods. As a result, there is at least a quantification problem in the assessment of quality in the public sector. Quality award competitions can be instruments measuring quality by using complex multi-dimensional indicators. This technical function is closely related to the third function of public quality award programs, which is to help public agencies improve their quality assurance systems by learning from each other. By quality competition awards, excellent public organizations are identified and their success factors are made visible to other agencies (Haubner, 1993). There is also a cooperative element in quality competition awards, which is perhaps the most important function of such awards if they claim to be an instrument in fostering innovation and quality improvement in the public sector.

It is obvious that there is a tension between the competitive and cooperative elements of public quality award competitions. On the one hand, participants of award programs want to know how good they are compared to other
organizations and to conduct, subsequently, SWOT analyses. On the other hand, nobody wants to “lose”. Organizers of quality awards have to stress the cooperative element of awards; namely that every participant “wins” by learning from others and that winning the award is not terribly important. This trade-off has to be made by organizers of award programs. If the learning aspect is central, it might be more functional for the program to avoid the term “competition” and to look for another term. On the other hand, if the aim is to introduce competition, winning an award becomes the crucial element of the program and “losing” has to be part of the outcome.

Like the Citizen Charter Award, quality awards only consider “good” organizations. “Bad” organizations are not considered. In this sense, the term “quality award competition” is somewhat misleading because quality awards are not a substitute for market competition since, officially, there are no losers. As Halachmi (1995) has pointed out, quality awards do not automatically have a motivational function for the “winning” organizations. For organizations which do not achieve an award, such competitions can be highly dysfunctional. Quality awards tend to be self-selecting in that organizations that do not even dare to participate in an award program are ignored. Does the award program have indirect motivational effects by giving such organizations an example? It becomes clear that the capability of such organizations to become “learning organizations” (Senge, 1990a; Argyris, 1992) is the precondition for benchmarking. Quality awards do not reach, and can have no impact on, those organizations which lack elements of a learning culture. Pessimistically, this means that quality awards are more likely to be a means of making relatively “good” organizations “better” and that they cannot “mobilize” the mass of public organizations and agencies. Nevertheless, quality awards have the important function of raising the awareness of quality in the public sector and for facilitating research on innovation.

The basic concept of benchmarking, however, raises further questions. First, does benchmarking stimulate innovation? It might be more convenient to look to what others are doing rather than think about original solutions. Benchmarking helps to spread innovation but its link to simulating innovation is an unanswered question. Second, which benchmark instrument best fulfils the performance measurement needs of public organizations? Attention needs to be given to the fact that benchmarking is a complex measurement instrument with high transaction costs. Thus, it needs to be ensured that measurement costs do not exceed organizational “slack” that is being eliminated in the name of economic efficiency (Kouzmin et al., 1996). Finally, the ultimate goal of benchmarking is to make organizations “lean and mean”. If, however, organizational slack is a precondition for innovation and crisis response capabilities (Rosenthal and Kouzmin, 1996) and benchmarking aims at reducing slack, benchmarking could be counter-productive, especially in industries, agencies and firms for whom innovation is a constituent element of competitive advantage.
From benchmarking to learning strategies

Recent European experience with benchmarking indicates that the benchmarking community is cultivating colourful differentiation; so much so that benchmarking can be now viewed as a terminological “umbrella” covering a wider range of processes than the core-elements conventionally viewed. Recent forms of benchmarking now include comparisons of organizations on the basis of self-assessment, but using identical indicators; benchmarking cooperatives – groupings of organizations or sub-units that get together in order to learn from one another; and, finally, benchmarking has broadened to incorporate empirical studies of the different paths by which social modernization processes take place. The development of instruments for self-assessment of public organizations and measurement of the degrees and profiles of specific modernization and reform strategies is also proceeding.

The creation of institutions called “innovation-rings”, in which change agents in several branches of the German public sector, for example, get together in order to learn from each other, is also a recent benchmarking innovation in Europe. Finally, some detailed work has emerged on the comparison of the performance of public sector organizations by means of employee, citizen or client surveys, the results of which are retrieved and relayed to informants by means of data-feedback techniques.

A preliminary attempt at drawing some general conclusions about expanding German and other European benchmarking experiences involves nine emergent propositions:

*P1*: The different forms of benchmarking have many common effects, although they differ with respect to certain other effects. As a consequence, one has been able to successfully practice combinations of different forms of benchmarking which have been successfully used with more differentiated functional profiles. This has been done without having a “clear-cut” theory at hand.

*P2*: Benchmarking can, on the whole, be seen as a learning strategy; as a widely-shared conviction and aspiration within the benchmarking community itself that having experimented with benchmarking, the learning potentialities and outcomes of benchmarking can be considerable and real. Empirical evidence supports this learning conviction and aspiration. For example, to a considerable degree, by merely sending public agencies award-questionnaires, many such organizations can be influenced by being forced to evaluate previous accomplishments and on-going weaknesses.

*P3*: A full utilization of the “learning” potential of benchmarking seems to be only possible if such forms of benchmarking are chosen or developed so as to combine elements of competition and cooperation. This proposition appears to be paradoxical in view of the traditional arguments that these two social mechanisms are antagonistic. One might call this the “dialectical” dimension to modern complexity. The
character of social and organizational learning itself might be understood better if conceptualized as a “dialectical” combination of both competition and cooperation.

P4: Recent developments in the field of benchmarking indicate that this “dialectical” combination of competition and cooperation is already being widely practised. This is quite apparent wherever “benchmarking cooperatives” have been created. The same impression prevails, however, if one looks more closely at the development of quality award competitions. As Halachmi (1995) points out, one danger inherent in award competitions seems to be the effects of frustration and “demotivation” that emerge whenever award competitors “fail” to achieve some recognition. Even “winners” can be afflicted by frustration and demotivation, however, when they reflect on the time and money associated with being a “role model” for other interested agencies or being the object of attention of less successful competitors’ envy, distrust, even their own employees’ heightened sets of expectations. In light of such experiences, the organizers of the European Quality Award, for instance, see to it that every participant receives a “reward” in terms of learning outcomes and that successful competing organizations take a cooperative role in this context. In the German experience, for example, it has come to be regarded as appropriate to combine competition and cooperation and this combination is now a decisive success factor of quality award competitions. Everybody wins by having a chance to cooperate and by taking a role as a learner or as an action-learning agent. However, some problems with the implementation of this ambition exist and a stable agenda for future action remains yet to be identified.

P5: The learning-effects of benchmarking are, to a very high degree, dependent on adequate organizational conditions and managerial solutions. This holds true for benchmarking activities themselves and only if the delicate relationship between competition and cooperation is cogently managed. The success of innovation rings, for example, is primarily dependent on the availability of such organizational and managerial skill. Such rings may either collapse under the pressure of mutual jealousy or be transformed into managerial complacency with a supportive, but non-learning, climate. Difficult problems of keeping a productive learning-oriented balance between competition and cooperation may also arise when the results of employee surveys are departmentally categorized and presented in a comparative way.

On the one hand, the benchmarking opportunity might be quite eagerly taken by actors inside the organization. On the other hand, actor satisfaction with the results of benchmarking surveys will strongly correlate with individual assessment scores. The effect may well be that neither the high scoring nor the low scoring units will be ready to take
on a learning role. The “winners” may just feel justified in their self-confidence, whereas the “losers” may feel highly defensive. Instead of enforced learning, a heightened level of organizational conflict can result from benchmarking if organizational or managerial skills are lacking regarding the need to harmonise “winners” and “losers” on the basis of a common commitment to learning.

**P6:** In many cases of benchmarking, mere organizational or managerial skills will not be enough. The so-called “upward” assessment of superiors is possible only where a cooperative atmosphere or culture has been previously established. Benchmarking will only be a reliable learning strategy if the precarious balance of competition and cooperation is supported by a favourable framework of contextual conditions as well. Comparative assessments of people and sub-units inside organizations cannot lead to productive learning behaviour if the organization is highly centralized; one in which managers are rigorously engaged in strongly competitive “face-saving” and employees are regarded as “inferiors”. The same holds true for comparisons between organizations, dominated by officials who think and act mainly in terms of achieving or defending a hierarchically defined status position.

Status-seeking behaviour, as such, is not a hindrance to benchmarking and to learning. On the contrary, one finds an element of this behaviour in nearly every “learning” activity. A serious impediment to learning occurs, however, wherever the status-seeking behaviour of some constrains the status-seeking behaviour of others. In hierarchical organizations, in which “vertical thinking” (Kouzmin, 1980a; 1980b; 1983; Korac-Boisvert and Kouzmin, 1995) prevails, this will necessarily be the case to a very high degree. The conditions for successful, learning-oriented benchmarking strategies are most effective, it is postulated, where the structural and cultural characteristics of an organization permit status-seeking behaviour by everyone. The main structural pre-conditions for this can be found in decentralization, the empowering of sub-units and the use of non-hierarchical means for coordination. In view of the state of debate in organization theory (Weick, 1969; Galbraith, 1977; Argyris and Schon, 1978; Bachrach, 1989; Meyer and Zucker, 1989; Masachiro *et al.*, 1990; Powell, 1990; Senge, 1990b; March, 1991; Davidow and Malone, 1992; Harrow and Wilcocks, 1992; Kotter and Heskett, 1992; Nohira and Eccles, 1992; Schmidt, 1993) one can formulate the paradox that the outstanding pre-condition for learning in organizations is the creation of the “learning” organization in structural and cultural terms.

**P7:** Dependency on contextual conditions implies the necessity to adapt benchmarking instruments to prevailing conceptions of “modernization”, for instance, as well as to given levels of actual development. The concept of “modernization of the public sector” is
culturally relative and the design of measurement instruments for quality award competitions, for instance, needs to be developed with such considerations very much in mind. The need to consciously define modernization targets, regarded as realistic under given conditions and compatible with elements of the international “new public management” movement used as models, is challenging.

P8: To the casual observer, benchmarking may appear as an element of little importance within the complex framework of conditions necessary for learning. Indeed, it would be an over-estimation of benchmarking if it were elevated and regarded as the “primary” success factor in public sector reform. On the other hand, benchmarking is a “necessary” element for such public sector modernization. Benchmarking, it needs to be stressed, has to be seen as a substantial aspect of nearly every activity of public sector reform, even where it has not been developed into a separate activity.

P9: The remaining question involves what might be done to foster benchmarking? Looking at the present state of public sector reform, one might easily argue that the potential of benchmarking is by no means exhausted. For instance, it is difficult to motivate public sectors to use self-assessment workshops. The use of indicators as means of comparative self-assessment is spreading, but the quality of these indicators is still rather variable. Furthermore, applications for quality awards at local, state and federal levels are less than what one would have anticipated as being worthy of successful consideration. Similarly, the creation of benchmarking cooperatives lags significantly behind the potential level of activities. Furthermore, there are considerable European differences in experiences to date with British, French or Belgian citizens’ charters; the number of employee-, citizen- and client-surveys is still very small and those surveys which have been done often lend themselves to limited implementation, if any. Finally, empirical studies of the different ways to modernize and implement reform, and to possibly become “excellent”, are scarcely done (Klages and Haubner, 1995).

Taken together, benchmarking is still a rather exceptional activity in many public sectors. The question of how to foster its use seems to be an urgent priority. The situation in the UK, for instance, is much more developed. The UK experience differs from Germany and other European public sectors by having established a citizens’ charter; by demanding performance indicators from local and state authorities; by having charged audit institutions with respective controlling functions; and by having created additional bodies such as the Citizens’ Charter Unit. To some extent, Britain might be regarded as a role model for other countries. On the other hand, the quality of those performance indicators used in the UK are subject to ongoing discussion and review (Pollit, 1990; Carter et al., 1992; Sanderson, 1992). A concluding message from
preliminary survey of European benchmarking and performance measurement experiences would be the call to learn from one another as much as possible; maybe with a motivation to imitate.

Benchmarking organizational learning through self-assessment

Given the widespread assumptions regarding the putative superior quality of private sector service delivery (Poister and Henry, 1994), self-assessment could also be helpful in placing attitudes towards government and public services into perspective, as well as providing incentives for improving current practices. For example “accuracy” and “reliability”, two core components of service quality, are not regarded as high for public sector products and services (Poister and Henry, 1994) as for those of the private sector. Furthermore, “responsiveness” and “sensitivity”, two other dimensions of service quality considered lacking from public programmes (Poister and Henry, 1994), can also be benchmarked. However, while direct comparisons such as benchmarking should be made where possible (for example, services provided by public health clinics and private health clinics, or private and public employment agencies), services such as social security benefits need to be self-assessed or compared with other government services such as students’ assistance benefits. To facilitate adequate benchmarking or genuine self-assessment, a “meaningful” corporate financial information system (CFIM) must be adopted by public sector organizations to provide a basis for benchmarking (Australia, Department of Industrial Relations, 1992). This need for a CFIM system is currently receiving attention within the Australian public sector.

For example, the new Australian Accounting Standard (AAS 29) required government agencies to adopt an accrual accounting system from 1995 onwards. However, it did not require the wholesale adoption of commercial accounting practice (Australia, Public Sector Accounting Standards Board (PSASB), 1993). Although the AAS 29 focuses on the cost of service delivery and direct cost recoveries, not bottom-line profit, it requires public sector organizations to prepare general purpose financial reports that disclose the resources they control (assets); the obligations they incur (liabilities); the financial characteristics of their operations during the period (the cost of service provided, cost recoveries and other revenues); and their major source of use of cash (PSASB, 1993). Accordingly, this information allows informed decision making and assists in ensuring that departments are accountable to the ultimate provider of funds and to the beneficiaries of services provided; taxpayers, customers and businesses. Furthermore, it can be used to facilitate the benchmarking of costs of public sector goods and services with industry parallels. This is a significant improvement considering that in recent history many public sector organizations did not make financial information publicly available other than through budget papers (Micallef, 1994).

Client/customer or citizen surveys are useful tools for evaluating public sector goods and services (Webb and Hatry, 1973; Fitzgerald and Durant, 1980; Streib, 1990; Miller and Miller, 1991; Watson et al., 1991) on a variety of criteria,
such as effectiveness (Usher and Cornia, 1981) and impact (Brudney and England, 1982) as well as quality and productivity (Folz and Lyons, 1986). However, client evaluations of the specific services they receive tend to be more positive than the rating of these same services by the general public (Katz et al., 1975; Poister and Henry, 1994). This influence may be overcome by a random sampling methodology. Furthermore, citizen responses to the very general satisfaction items in citizen surveys often do not correlate with more direct, objective measures of service delivery (Stipak, 1979). Linking survey data with more objective indicators (Ostrom, 1982; Brown and Coulter, 1983; Parks, 1984) such as benchmarking can aid organizational learning processes, agenda setting and quality improvement through the balanced assessment of programme performance. Customer/citizen survey data can be especially helpful in agenda setting in seven broad areas: customer expectation; work culture; work design; work-force requirements; hours of operation; costs; remuneration and evaluation (Graham, 1994). The effects of these agenda items can be successfully benchmarked.

Competence can be benchmarked on current skills, multi-skilling, skills attainment, future needs, empowerment, staff distribution, team-work and commitment. The cost of the delivered product or service can be benchmarked (or measured against a quality standard) on programme costs, staff costs and delivery costs. For example, the effects of a set agenda for cultural change through the empowerment of employees (the product is cultural change) can be benchmarked against external or industry standards or, in their absence, evaluated against internal standards based on the following items (Wilson and Durant, 1994, p. 144):

- the extent to which decisions have been decentralized to lower organizational levels;
- openness, closeness or the extent of hierarchical decision making (who and how many actors have freedom to decide?);
- availability of opportunities for avoiding responsibility (the opportunity for “flight”);
- the extent to which organizational resources have been distributed to employees;
- the extent to which a multi-level commitment exists toward change;
- how well institutional incentive systems reinforce, rather than discourage, cross-functional cooperation (the project team linkage); and
- the degree of integration of functional units that contribute to the production of goods or services (the patterns of structural independence).

Empowerment demands continuous learning and risk-taking activities and both thrive in a climate of trust and confidence, while both wither in a climate of fear (Holpp, 1994). The fundamental challenge of empowering is the
requirement that organizational actors continuously change and learn. The activities that should be benchmarked are learning activities and change as a consequence of learning. From the behaviourist perspective, this can be achieved by “successive approximations”, or the series of steps that bring actors closer to a fully learning organization by gradually introducing them to new tasks and procedures, helping them to learn different jobs, reinforcing changes in their behaviour and making sure that they are provided with a safety-net adequate for different learning speeds and styles. Considering that empowerment, like learning, is an incremental process; that empowerment in most cases requires culture-change programmes that take a minimum of 9-12 months and often up to 18 months (Brooks, 1994); and that established a priori standards are often absent (Sims, 1992), the effects of change may be benchmarked in stages (Wilson and Durant, 1994, p. 144):

- stages of employee readiness level (level of skills, resources and commitment necessary for change);
- stages in the ability of employees to cope psychologically with the change sought in long-standing behaviour (the extent of behavioural change required); and
- stages in the ability of groups to accommodate change (the actors’ success rate in broadening their jobs; bottom-line results delivered in terms of costs, quality and productivity).

Employee readiness levels or attitudes towards change may be measured through interviews and surveys. Broadening job scope or job enrichment may be measured through job analysis, direct observation and measures of actual job accomplishment that then can be benchmarked against industry or internal standards (Holpp, 1994). Achieved productivity can be gauged when fewer resources are used to achieve more while the quality of the service or product remains as high or becomes higher. The other measure that can be used to gauge learning is that the number of significant administrative failures, or “soft-core” disasters, are less over the time than the industry or internal standard (Kouzmin and Korac-Boisvert, 1995). This is particularly relevant in the area of IT development where past and current failure rates are approximately 80 per cent (Legge, 1988; Dampney and Andrews, 1989; Korac-Boisvert and Kouzmin, 1994) and with a predicted future failure rate of 50 per cent, with a 0.7 probability, due to organizational resistance to change (White, 1993). Gartner Group research suggests that high rates of failure will persist in the 1990s with leading-edge technology, such as client/server and distributed processing, but with a larger variety of failures (Gartner Group, 1992; 1993; Thompson, 1993). Various software matrices can also be useful benchmarking tools in the area of IT development.

These stages must be well documented so that they can be used as comparative instruments. Thus, the documentation of each pre-defined stage of change serves not to entrench a process but to measure it and improve on it.
The culture of continuous improvement is one of ongoing re-examination and renewal of the format of work (Graham, 1994). In continuous improvement, the perceived “best practices” or industrial awards, job descriptions, employment categories and promotion positions change to meet client demand (Graham, 1994). Conditions of employment are necessarily flexible rather than rigidly demarcated. Furthermore, changing culture to one of learning through empowerment and self-assessment often involves various re-training initiatives (Talley, 1991; Argyris, 1992) which, owing to the pragmatics involved, are often in stages. Notwithstanding, amongst the greatest threats to cultural change through training exercises are organizational defences rooted in bureaucratic routines and interests that are stubbornly resistant to change (Argyris, 1992).

Other criteria can be used for evaluating quality efforts, such as the private sector’s quality evaluation approach: the Malcolm Baldrige National Quality Award, which consists of seven criterion categories – leadership; information/analysis; strategic quality planning; human resource utilization; quality assurance of products and services; quality results; and consumer satisfaction (US National Institute of Standards and Technology, 1991). Premised on the Baldrige Award, the eight-category criteria for the public sector, developed by the Federal Quality Improvement Prototype Award, substitutes the human resource utilization item with two separate items: training/recognition and employee empowerment/team-work, each with a different weighting score (US Federal Quality Institute (FQI), 1992).

**Conclusion**

A review of job descriptions, performance agreements, performance appraisals’ instructions and empirical research indicate that public sector organizations still promote actors with traditional views (James, 1992). Those who have the ability to set direction, make key decisions, energize and control staff, only rise to the fore in times of crisis (Senge, 1990a; James, 1992). The empirical evidence suggests that public sector organizations have a tendency towards high absenteeism and low turnover rates (Wooden, 1990; Callus et al., 1991). Values are rooted in an individualistic, localized culture and reinforce a focus on short-term events and group heroes instead of systemic forces and collective learning. This evidence suggests an emergent need for organizational cultural change, where leading actors become designers, teachers and stewards (Senge, 1990a). They are responsible for building organizations where people are continually expanding capabilities for learning (Senge, 1990b).

The process of periodic public sector reform followed by a significant period of consolidation is based on the premise that organizational homeostasis is the norm and that reform is the exception (Graham, 1994). This assumption needs to be replaced by one of continuous improvement, where the ongoing re-examination and renewal of the work context is the norm. This, in turn, requires the construction of an atmosphere in which participating managers and other stakeholders alike feel sufficiently free from threats to their self-
esteem to be “open to learn”. Helping actors move towards becoming more effective and bringing about positive change in their own behaviour is not sufficient; it is equally important that they understand the relevant processes of personality development, leadership and group dynamics (Davis, 1994). Benchmarking changes which are consequences of learning can facilitate actor learning and understanding of evolving organizational changes.

Notwithstanding that existing technologies are becoming better focused (customer feedback, prototyping), more disciplined (re-use, design reviews, quality matrices) and that new technologies (object-oriented tools, automated software synthesis) have emerged, a significant challenge for IT management in the public sector, for example, remains. The transfer of software technologies into organizations and bridging the gap between state-of-the-art and the state-of-practice remains problematic. Concerted efforts by both researchers and practitioners are required to engineer a bridge for the chasm. In Australia, for example, over 52 per cent of IT professional services (consulting; technical support; system integration; software development; outsourcing of training; development and information processing) are in the public sector (IDC, 1992) and, as such, beg improvement and learning from past disasters.

The use of quality models and other forms of structured experience in software development, along with feedback and continuous learning, need to become a genuinely adopted practice and not be paid lip service or fobbed off as a passing trend. Client services are becoming increasingly important forces for government, industry and the wider community. In order to harness an organization’s creative potential, ensuring that its actors are well equipped to exploit the increasingly important role software development is playing in society, there is a need for “learning to learn” (Argyris, 1982) that produces continuous improvement. Considering that most software/applications development are group activities, involving all the complexities of group dynamics such as communication, networks and organizational politics, there is a need to re-examine the human and organizational sides of software development. The social, psychological and cognitive factors in software development are all related to quality improvement failures in organizational settings, particularly in the problem definition stage (specification requirements and design).

Public sector organizations are becoming not only users, but providers and exporters of global information and associated services in an increasingly globalized market. Benchmarked knowledge and information are the fundamental strategic resources of the age; access to them through electronic networks is vital to overcome the numerous “tyrannies of distance” afflicting the many domains of public sector organized action.

References


Gartner Group (1992), *From Computing to Networking Information Technology: The Next Five Years (Executive Summary Report)*, Gartner Group, Stamford, CT, September.


Learning for agency effectiveness


Further reading