
4. Competing and complementary approaches to evaluation

Reinhard Stockmann

The aim of this chapter consists in providing information about competing and complementary approaches to evaluation. Before that (Chapter 4.1), a look is taken at the institutionalisation of evaluation. It shows that evaluation has massively gained in importance worldwide in recent decades. However, at the same time it becomes clear that evaluation is facing some major challenges. One of those is associated with the range of alternative concepts and approaches available, some of which even integrate evaluative elements. That means that this discipline, still young in professional terms and still somewhat insecure in institutional ones, may come under pressure, because it could be displaced by older approaches which are already well established. So it is that one sometimes hears the hypothesis put forward that the best days of evaluation are over and that auditing is ‘the new evaluation’ of the future (Pierre, Peters and de Fine Licht 2018, p. 736). Chapter 4.1 shows that that is not in fact the case, but nevertheless, the challenges posed by alternative concepts to evaluation can hardly be overlooked. For that reason, but not only for that reason, it is important for evaluators to know about these competing and complementary approaches: on the one hand, to be familiar with the competitors and their performance profiles and overlaps with evaluation, and on the other, to use elements of those concepts to investigate combined deployment options and sharpen their own (evaluative) profile.

Chapter 4.2 therefore presents the concept of controlling, which has a number of things in common with monitoring (Chapter 3). Auditing (Chapter 4.3) has expanded a good deal in recent years, which led Michael Power as far back as 1999 to title his book *The Audit Society*. Furthermore, auditing can be seen ‘poaching’ more and more in the territory of evaluation and competing with it, above all in the form of the performance audit.

Another method that can be combined well with evaluation is benchmarking (Chapter 4.4). Benchmarking aims to enable performance to be compared so as to use existing solutions to problems for the improvement of one’s own products, services or processes. The data, analyses and assessments necessary for this can be procured via evaluations.

Chapter 4.5 ventures a synoptic comparison of the concepts dealt with here with monitoring and evaluation, illustrating the various performance potentials, common features and differences.

In the final section, Chapter 4.6, the concepts and instruments presented are integrated into the context of overarching quality management concepts such as total quality management (TQM), ISO certification and new public management (NPM). It becomes clear that the existing management instruments have so far not been combined much at all, and considerable amounts of potential for innovation and learning are thus being wasted. The chapter concludes with the theory that with a combination of KPI-based instruments, approaches for ongoing performance recording and evaluations which have their strengths in impact-related and cause-

related analyses, the dirigibility of programmes and organisations and their overall quality management could be improved considerably.

4.1 INSTITUTIONALISATION OF EVALUATION

In recent decades, evaluation has taken a tremendous upturn worldwide. In politics, society and science, it is deployed for the objectives described in Chapter 2 – with varying intensity. Having said that, its *institutional anchorage*, which ensures a certain stability and continuity in its application and use, has a wide range of manifestations in the three areas and in individual countries. This is shown by the few studies that exist on this topic (cf. Furubo, Rist and Sandahl 2002; Rosenstein 2013; Jacob, Speer and Furubo 2015; Stockmann and Meyer 2017a). A large-scale research project which has set itself the target of recording the institutionalisation of evaluation in all the relevant countries in Europe, the two Americas, Asia-Pacific and Africa, currently offers the most comprehensive insight. The two volumes that have appeared so far on Europe and North and South America (cf. Stockmann, Meyer and Taube 2020b; Stockmann and Meyer 2021a, 2021b, 2021c; Stockmann, Meyer and Szentmarjay 2022) show clearly that in spite of the enormous cultural, social, economic and political differences within these continents, and of course even more so between them, there are *more things in common than differences*. The most striking thing is that, on the one hand, it can be said in general for all the countries investigated that evaluation is most firmly anchored in the political system, but a closer look reveals some major differences between the continents: when the legislative bases are compared, it becomes clear that in Europe evaluation is only prescribed in national laws in 3 out of the 16 countries investigated, and even national (cross-sectoral) ordinances exist only in approximately half of them. In the Americas, by contrast, there are national evaluation laws in 8 countries out of 11 and all of them have national evaluation ordinances.

With regard to the role of evaluation in the national audit offices too, there is a serious difference to be seen: whilst in Europe the audit offices explicitly also perform evaluations in approximately half the countries, there are only two that do so in the Americas: the USA and Colombia.

If we analyse the *users' side*, we get a completely different picture. Measured against the sectoral propagation of evaluation – the number of policy fields in which evaluations take place on a regular basis – and against the intensity and frequency at which evaluations are conducted, Latin America is seen to have a *major implementation deficit*. In other words, although in Europe, viewed overall, there are far fewer legal and organisational evaluation structures, the degree of use is considerably higher in Europe than it is in Latin America.

When it comes to the commissioning and use of evaluations by national *parliaments*, by contrast, the picture is fairly uniform: in all the countries examined, parliamentarians have, as a rule, very little interest in evaluations and their results.

When the indicators on the legislative and organisational anchorage of evaluation are combined into a dimension referred to as ‘institutionalisation’ and correlated with the cumulated indicators for the dimension ‘utilisation of evaluations’, this reveals only an average correlation. However, it also becomes clear ‘that whilst legislative and organisational anchoring supports the extent to which evaluation is used in the political system, *this institutionalisation is far from guaranteeing its use*. To put it another way, *there is no country in which evalua-*

tions are carried out to a large extent in spite of the fact that there is no legal framework' (cf. Stockmann, Meyer and Szentmarjay 2022, p. 512).¹

The findings on the prevalence of evaluation in the *social system* of a society reveal a major similarity. For all the countries investigated in America and Europe it can be said that the communalities are surprising: the contribution of civil society for developing evaluation in the countries is not influenced by the size, degree of organisation, financial capacities, political influence and engagement, or by the philosophy, spirit or cultural diversity either; it is close to zero everywhere (ibid.). It tends to be the state that imposes rules for evaluation on civil society organisations, particularly when it is funding them. Only in a small number of civil society organisations – particularly in the sectors of education, social services and development cooperation – is evaluation used as a management instrument.

The demand for evaluation, which is determined above all by the political and social system in a society, is opposite the supply side. Here, it is above all the universities and institutions of tertiary education which play an important role with the facilities for education and training they offer. One astonishing thing is that the demand is almost completely decoupled from the supply side. That means that there are a number of countries in Europe and Latin America that offer educational programmes, although there is practically no demand for evaluation in the political and social system – and vice versa! From that, the authors draw the following conclusion: 'The institutionalisation of evaluation in the academic system is remarkably poorly linked to the institutionalisation in the political and the social system – and therefore to the formation of an evaluation market and the forces of demand and supply' (Stockmann, Meyer and Szentmarjay 2022, p. 514).

One thing that the countries investigated in America and Europe have in common is that almost all of them have *voluntary organisations for professional evaluation (VOPEs)* in the form of associations or at least casual networks. Having said that, these are not well known in society, have but little organisational power, and only a slight tendency to become active as lobbyists for evaluation, either socially or politically. Alongside education and training and organisational constitution, *academic standards* have an important part to play in the professionalisation of a discipline. Here, the authors' analysis of countries and continents revealed across the board that 'there is still a lack of obligatory standards for improving the quality of evaluation and to increase utilisation both within organisations and within societies as a whole' (Stockmann, Meyer and Szentmarjay 2022, p. 515).

Overall, it can be said that in recent decades evaluation has made tremendous progress in American and European countries, even if that is not true of all the countries to the same extent. That progress did not develop evenly, but in waves, which began in the USA in the 1950s and 1960s. Meanwhile, some of the so-called 'frontrunner countries' (in the Americas: USA, Canada, Mexico; in Europe: Great Britain, Denmark, Germany) have lost momentum, whilst others have graduated to the leading group, some of them in an astonishingly short time (measured by the degree of institutionalisation of evaluation in the three systems analysed).

The question of what it is that has *promoted* the *institutionalisation* of evaluation in the Americas and in Europe in the last few decades is one that, at least for the political system, is surprisingly easy to answer: in spite of all the international and continental differences, there is a clearly identifiable force, namely the *efforts made to modernise and rationalise the admin-*

¹ Emphasis by the author.

istration. The reforms associated with that had at least two roots: financial administration and NPM (or output / outcome / results-based management).

The origins, which lay in financial administration, are associated with the intention of improving the *financial control* of the state. Planning, programming and budgeting systems (PPBS) were developed to that end. These systems were a variant of the management approaches popular at the time, which aimed to improve the effectiveness and efficiency of systems and the allocation of funds by linking explicitly defined organisational targets with output and impact parameters.

The second source of reform comes from efforts made to improve job performance and the process of rendering services in public administration. To achieve that aim, the concepts of NPM, output and outcome-oriented and results-based management were developed.

The most significant thing for evaluation was the associated departure from the centuries-old *management of administration via input parameters*. The way that worked was that the administration was encouraged, by the allotment of various-sized inputs (such as financial resources, personnel, equipment) to become active in certain assigned areas of work (cf. Stockmann 2008, p. 58). For the *new management models*, which require output, outcome and impact-related data, the traditional instruments of administration are not enough; other techniques and instruments are necessary. This was the breakthrough for evaluation, which provides methods for the procurement, analysis and assessment of such data. Thus it is not an exaggeration to say that: ‘Without this kind of administrative modernisation and the concept of making strategic political and administrative decisions based on a rational informational underpinning, evaluation would have remained insignificant’ (Stockmann, Meyer and Szentmarjay 2022, p. 518).

These findings apply to the institutionalisation of evaluation in Europe and to that in the Americas, with one major *difference* which has continued to have consequences right up to the present day: in the Americas, evaluation aims to *rationalise budget-relevant decision-making* much more strongly than it does in Europe. Above all, it has been placed in the service of the financial administration and thus first and foremost in the service of financial control, budget allocations and accounting. That means that evaluation, like audit, is perceived and used predominantly as an instrument of control.

In Europe the development of evaluation was not so closely linked to this tradition or to necessities of financial and budget administration, but more strongly to the concept of NPM and results-oriented management.

That enabled other performance criteria such as *goal achievement* and *output and outcome investigation* to come to the fore. Because of this control image, in many countries – particularly in Latin America, but also in some European countries – evaluation is still perceived today as a threat and not as an instrument for improving the efficiency and effectiveness of political or administrative measures and regulations.

Another *driving force* in the development of evaluation in Europe is the European Union. By making the conducting of evaluations one of the obligatory conditions for the receipt of support from the European Structural and Investment Funds (ESF), the EU has ensured that evaluation is also practised in countries which had not applied the concept before.

In Latin America there are no comparable actors, like the EU, that would have been able to bring such a formative force to bear on the development of evaluation. A role that, at best, supports that development can be attributed to the national and international donor organisations (cf. *ibid.*).

In spite of the deficits that can be seen, the development of the institutionalisation of evaluation in the last few decades is to be assessed as a success story. So that this story can continue and the gaps diagnosed can be filled, evaluation will have to cope with a few *challenges* in the future. These include developments which can be seen in various countries in Europe and the Americas:

- The increasing standardisation of evaluation involves the risk of *bureaucratic routinisation*, which debases it to become a technology adapted to suit the administration.
- If evaluation is seen above all as an instrument of control from which negative sanctions are to be expected, it is perceived as a *threat* that should be avoided.
- Most of the VOPEs for evaluation see themselves primarily as professional associations that unite evaluation practitioners and scientists under a single umbrella. As they tend to have weak organisational structures and do not consider lobby work in society and politics to be their job, a *central promotor* for the dissemination and application of evaluation is lacking.
- In many countries, the *degree of professionalisation* of evaluation is still low: its *education and training* is not firmly established in institutional terms. In almost all these countries there is a tremendous fluctuation on the supply side. *Evaluation standards* that are well known in the public arena are often lacking, and a *certification of evaluators* is only possible in very few countries. Thus evaluation is not a profession of the kind protected by law. Anyone who feels competent, and has the many years of evaluation experience which are very often cited in invitations to tender as a central criterion for selection, can tender for and conduct evaluations. That leads to very *variable evaluation qualities*.
- In addition, evaluation is, as it were, in *competition* with other instruments used in management decision-making. Apart from that, it can be seen in individual countries that evaluation is blended with instruments such as quality management, organisational consulting, controlling, audit etc. This blending is causing the *distinct profile* of the ‘core trademark’ of evaluation to be lost, a profile that involves gathering data, analysing them independently and in accordance with scientific standards, and assessing them with the aid of empirical methods and transparently defined criteria. In such a case, there is a danger that evaluation may become a replaceable tool that no longer has any character of its own.

For several reasons, it thus makes good sense for an evaluator to occupy himself with these ‘other’ instruments or approaches (Chapters 4.2–4.4). This is not for the primary purpose of setting himself aside from them, isolating himself or perhaps even erecting a wall to keep them out, but so as to be able to explore complementary, ancillary or combined options for deployment without injuring his own professional standards. So evaluators should know what those ‘other instruments’ are, what they are used for, what concepts they are based on, and what similarities and differences they have when compared with evaluation.

Our aim here is to present some of these approaches and instruments. The emphasis is on the presentation of the basic conceptional factors and not on the application. The latter would go beyond the scope of this chapter, which concludes by mapping out the similarities and differences between the various approaches as compared with evaluation (Chapter 4.5) and illustrating their various fields of application (Chapter 4.6). It thus becomes clear that the various different concepts can be used complementarily.

4.2 CONTROLLING

4.2.1 General Introduction

Controlling² developed from the field of *accounting*. The ‘confusing diversity’ (Horváth 2009, p. 127) of controlling concepts makes it more difficult to arrive at a generally acceptable description (cf. Becker and Baltzer 2009). Looked at in functional terms, most of these concepts³ see in controlling the *subsystem of corporate management* which ‘coordinates planning and control and the supply of information in a way that forms and couples systems in a target-oriented way, thus supporting the adaptation and coordination of the system as a whole’ (Horváth 2020, p. 62).

Controlling concepts are often based on *system theory considerations* (cf. Eschenbach 1996, p. 45ff.; Habersam 1997; Eschenbach 1999, pp. 8, 113f.; Horváth 2002, p. 98ff., 2020, p. 38; Jaspersen and Täschner 2012, p. 3; Reichmann, Kießler and Baumöl 2017, p. 56). The enterprise is viewed as a *social system* divided up into management and executive subsystems. As a social system, the enterprise is interactively connected with its environment. It is the job of management to react to the requirements of the environment in an appropriate way and actively shape them as far as possible. Management must correctly assess their complexity and dynamism and develop forward-looking strategies which safeguard the continued existence and prosperity of the enterprise. The more complex the environment, the greater the need for controlling induced by that environment. Given that, as a rule, the complexity of organising the enterprise thus increases with the complexity of the environment, the need for controlling induced by the enterprise increases too. Since it is assumed that the complexity and dynamism of the environment have increased markedly over time, management must react to the rapidly changing requirements fast and flexibly and improve its ability to innovate. In order to be able to do that, it requires *information*, the procurement of which is a central task of controlling: ‘Planning and control require a supply of information. It is a matter of providing all the information required for planning and control with the required degree of precision and agglomeration, in the right place and at the right time’ (Horváth 2020, p. 189) and as Eschenbach (1999, p. 9) adds, in order to ‘process it and forward it to the right recipients in a timely manner’ (cf. also Jaspersen and Täschner 2012, p. 431).

² Controlling is still a relatively young discipline. It is true that its origins are thought to lie in British royal court administration in the 19th century, but its genesis in enterprises in the USA only dates back to the end of the 19th century. University chairs with the explicit denomination ‘controlling’ were not established in Germany until the beginning of the 1970s (cf. Becker and Baltzer 2009, p. 2). For the history of the genesis of controlling cf. the concise overview by Gerlich 1999, p. 3ff. On the term controlling and its methodology cf. also Bähr 2002; Bethke 2003; Ehlers and Schenkel 2005; Spraul and Oeser 2007; Müller 2009, 2017, 2020; Czenskowsky, Schünemann and Zdrawomyslaw 2010; Piontek 2010; Baier 2013; Baum, Coenenberg and Günther 2013; Friedl 2013; Küpper et al. 2013; Jung 2014; Behringer 2021.

³ A good overview of the various controlling concepts is to be found in Becker and Baltzer 2009, p. 13ff. We do not aim to enter the discussion on the delineation between management and controlling here. Cf. instead Eschenbach and Niedermayr 1996, p. 71; Eschenbach 1999, p. 10; Jaspersen and Täschner 2012, p. 425ff.; Horváth 2020, p. 106f.

Controlling is not just supposed to ensure a supply of information, but also to work toward a situation in which decisions can be made and measures implemented on the basis of that information. By its transparency, controlling creates inherent necessities which increase the pressure to make decisions and act. The management must then take care to ensure that the decisions it has made are actually implemented in the enterprise.

Coordination is the second main function of controlling, ‘to get a grip on the complexity of corporate governance’ (Hoffmann, Niedermayr and Risak 1996, p. 48). Above all, controlling should ensure that the generation and provision of information are coordinated with the information requirements, and the planning and control system with the information supply system (cf. Horváth 2020, p. 63).

The extent to which a *controlling system* is institutionalised in an enterprise at all, not just functionally, but also in terms of its personnel, depends above all on its size: in a small enterprise, controlling, as a business management service function, is not separately institutionalised at all; the executives exercise the controlling functions themselves or have recourse to external providers.

In the literature there is a broad consensus to the effect that controlling supports management in its planning and control decisions. It has, as it were, a service function. For that reason, controlling is often compared with a ‘pilot service’ (cf. Eschenbach and Niedermayr 1996, p. 51; Gerlich 1999, p. 8; Deyhle and Hauser 2011, p. 6; Horváth 2020, p. 189; Mäder 2021, p. 31). Controlling gives management concrete assistance in strategy planning and strategy development, strategic control and strategic foresight, operative entrepreneurial planning and budgeting, and operative forecast (cf. Eschenbach 1999, p. 28; Behringer 2021, p. 3). Various instruments are deployed for this purpose. *Operative controlling* is geared above all to internal aspects of the enterprise and concerns itself with developments which are already manifesting themselves in the present via expense and yield. Thus there is an accent on the use of decision-oriented cost and yield accounts. Classically, operative controlling avails itself, above all, of the instruments of internal accounting, key performance indicators (KPIs), steering prices and budgets and the forecast (target-performance comparison).⁴

Strategic controlling, by contrast, is directed toward the future and gains importance with the strategic orientation of enterprises. Strategic controlling is intended to improve the ability of the enterprise to adapt and innovate by revealing exogenous and endogenous changes in the environment and encouraging the management to take timely action (cf. Habersam 1997, p. 97; Moos and Peters 2021, p. 43f.). The following instruments, among others, are deployed for this: analysis of strengths and weaknesses, analysis of sector and competition, analysis of potential and portfolio, analysis of cost structure, monitoring of implementation and results.

With regard to the *dissemination* of these instruments in German enterprises, empirical investigations make it clear that it is above all classical controlling instruments such as budget control, deviation analyses, reporting, cost accounting etc. that are deployed. Strategic instruments are used far less (cf. Niedermayr 1994, p. 144ff.; Baltzer 2013, p. 109ff.; Müller 2017, p. 138ff.) In addition, there is criticism to the effect that the latter are – in the great majority of cases – merely a further development of the operative instruments and that they lack theo-

⁴ A clear overview of which controlling instruments are used in which controlling concepts is provided by Becker and Baltzer (2009, p. 29ff.). In it, the focus on instruments of investment and cost accounting, cost management, deviation analyses and indicators becomes clear.

retical penetration and practical testing (cf. Küpper, Weber and Zünd 1990, p. 288; Langguth 1994, p. 1; Habersam 1997, p. 83; Weber and Schäffer 2020, p. 426f.; Mäder 2021, p. 116). Horváth describes the *critical aspects* of controlling as follows: ‘Excessive complexity due to a distinctive controlling hierarchy; heavy workload in operative areas because of information requirements; problems of acceptance because controllers are too strongly oriented toward control; and last but not least: poor supply of information (too slow, insufficiently based on recipients’ needs, false signals)’ (Horváth 2020, p. 447).

Another thing that is criticised about the controlling approach is the *one-sided image of humans* on which it is grounded, an image ‘that sees employees as resource-wasters and, accordingly, as employees who need to be supervised’ (Habersam 1997, p. 75). For that reason, the actions of controllers are ‘often characterised as being shaped by mistrust’ (Schäffer and Weber 2017, p. 9). The *understanding of the decision-makers’ role* which is prevalent in controlling is rejected as idealised and unrealistic, because its ‘deficiencies of will’ (Mäder and Hirsch 2017, p. 94) are not taken into account. Fault is also found with controlling inasmuch as it limits itself too unilaterally to the role of ensuring rational corporate governance, neglecting ‘value-rational’ issues (cf. Schäffer and Weber 2017, p. 11). Moreover, it is pointed out that controlling gives too little support to the entrepreneurial *error culture*, although that would be essential for the process of learning from mistakes of the kind it reveals (cf. Weber and Schäffer 1999, p. 731ff.; Haaker 2017, p. 35). Other frequently voiced points of criticism are the lack of *scientific clarity* of the approach and the *trust* controlling has in *planning*. Controlling is criticised as a *technocratic approach* which serves the sophisticated dirigibility of the organisation and its members (cf. Habersam 1997, p. 75f., 134; Bauer 2015, p. 40ff.). Other frequently expressed points of criticism relate to the problems involved in dissociation from the areas of management and corporate governance and the ‘arrogance’ of setting up what is more or less a kind of meta-management (Mäder and Hirsch 2017, p. 94; Mäder 2021, p. 122).

This lack of clarity leads to a situation in which controlling systems vary greatly depending on cultural contexts which, in enterprises that operate internationally, ‘lead to poor strategic decisions, permanent conflicts within the enterprise, unsuccessful cooperations and, in the worst case, economic failure’ (Funk and Rossmann 2017, p. 282).

A look at the *spread of controlling through sectors of the economy* shows that it is deployed above all in the production (industry, commerce) and service sectors (in particular financial services). Controlling is a management instrument or a leadership component, and is in principle not tied to any particular goals (e.g., making a profit), certain business activities (e.g., producing goods or rendering services) or sizes of organisation (cf. Küpper, Weber and Zünd 1990, p. 282; Habersam 1997, p. 96). Although it is assumed that controlling can be used in all organisational and social systems, controlling concepts were not implemented in non-profit organisations until the beginning of the 1990s (cf. Bauer 2015, p. 46). Whilst some NPOs use classical controlling approaches, which limit themselves to mere instruments of financial economics such as bookkeeping, cost accounting, economic planning and the reporting that goes with it, others also include qualitative parameters for service output and service provision targets with the aid of indicators (ibid.) In politics and administration, controlling – like monitoring and evaluation – has gained in importance thanks to the introduction of new management models (such as NPM, results-based management etc.) (cf. Chapter 4.1). Wollmann (2009, p. 2f.) ascribes a ‘strategic key function’ to controlling in the context of the

new management models, though he says that it is not yet being used adequately by politics or administration.

SUMMARY

- ✓ Controlling is a subsystem of management that provides information and relevant data on changes in the environment that have already taken place, and also possible future environmental changes, in order to create the prerequisites for adaptation and measures that influence the environment.
- ✓ Controlling supplies data from within the organisation to inform the decision-makers on a continual basis about the comparison between planned and actual developments – target-performance comparison – so as to make it possible for purposive corrections – management decisions – to be made.
- ✓ Controlling can also fulfil coordination tasks for the management system by creating the prerequisites for the coordination of the action of the individual subsystems in the management of the enterprise or, put more generally, of an organisation. Planning, control and information systems, for example, are established to this end.

4.2.2 A Comparison between Controlling and Monitoring

If we compare the tasks of controlling with those of monitoring (Chapter 3), we find a few common features but also some major differences (cf. Figure 4.1).

Criteria	Controlling	Monitoring
Historical origin	Accounting	Social sciences
Task	Non-assessive collection and coordination of information	Non-assessive collection of information
Addressed to	Management non-assessive	Management and company as a whole non-assessive
Reporting focus	Above all aspects of cost, in-house processes, no impact data, centred around input / output	Broad internal and external structures and processes, also data on intended and unintended effects
Frequency	Ongoing	Ongoing
Organisation	Top-down	Top-down or bottom-up
Criteria	Largely codified	Largely open
Methods	Especially cost accounting	Social science spectrum
Data sources	Exclusively quantitative	Quantitative and qualitative

Figure 4.1 Controlling and monitoring in comparison

Like controlling – but unlike evaluation – monitoring is an *ongoing task*, aiming to inform the management continually by means of data from inside and outside the organisation about the comparison between planned and actual developments (target-performance comparison), so that it – the management – can make purposive corrections. One thing the two methods have in common is that they supply information for decision-making but do not assess it. They contribute to the decision-making process but are not part of the decision itself.

Having said that, there are also a number of *differences* to be pointed out between these two procedures: an important one consists of the fact that monitoring only has an information supply function, whilst controlling (in some concepts) also has a coordination function. Another difference is to be seen in the fact that the reporting range in monitoring is far wider. Whilst controlling systems continue in practice to concentrate mainly on aspects of cost, this is an area often neglected in monitoring systems. But then a monitoring system also delivers data on the intended and unintended effects achieved.

Whilst controlling focuses very much on structural factors, monitoring also takes procedural and systemic questions into account. That prevents it from being reduced to a mere technical planning aid. Monitoring is necessary for the very reason that the planning and implementation of measures is understood as a permanent, ongoing process which must react to changing environmental conditions.

Employees and other stakeholders are often involved in the configuration and development of monitoring (cf. Chapters 6.2 and 7.2). Controlling, by contrast, is determined by fixed ‘controls’. Controlling follows a ‘top-down’ approach, whilst monitoring can be organised as a ‘bottom-up’ approach. Given that co-organisation is at least possible in monitoring but not in controlling, the criticism levelled at the latter – that of subtly manipulating employees – cannot be applied to monitoring.

Monitoring and evaluation are based on the same philosophy-of-science principles, and the same methods can be used for data collection and data analysis. These are far more comprehensive than the methods used in controlling, which are greatly reduced to cost accounting. The consequence of that is that in a monitoring process monetary or quantitative parameters are used alongside qualitative data, whilst in controlling they are used almost exclusively. Whilst controlling is a rather *rigid* method, monitoring is far more *flexible*. What is to be measured and how it is to be measured are things that can be specified in an interactive process.

Given that controlling originated from the business management field of science whilst monitoring originated from social science, i.e., they have come from different traditions, and that the two methods are as a rule used in different areas of application – enterprises vs. non-profit organisations / programmes – there have hardly been any attempts so far to unite these two instruments. Nevertheless, it appears that this might be possible. Monitoring approaches could contribute to enhancing controlling both in terms of its content and methodologically, ‘democratising’ it and opening it up to ‘subjective’ reality (i.e., to the various different perspectives of the stakeholders). By contrast, controlling could fecundate monitoring by including the aspects and methods of cost accounting.⁵

⁵ Habersam (1997, p. 186ff.) undertakes a comparison of the main determinants in the controlling and evaluation discussion. In it, however, he concentrates solely on the constructivist-empiricist approach of the ‘4th generation of evaluation’. In doing so, he fails to mention that this approach is by no means the only one advocated and accepted in evaluation research. To that extent, the comparison he makes paints a false – or very limited – picture. Other

4.3 AUDIT⁶

4.3.1 General Introduction

The audit, like controlling, developed from accounting. Originally its main *aim* was to check sets of financial figures according to accounting principles: ‘– that is, a retrospective examination for the purpose of forming an opinion of their fairness in conformity with generally accepted accounting principles’ (Owen 2021, p. 262).

Meanwhile, the term audit is used for a range of other activities. A distinction is made between the traditional *finance audit*, carried out retrospectively for the purpose of ascertaining the correctness, accuracy and integrity of financial data, and the *performance audit*. The latter is defined by the European Court of Auditors (2017, p. 7) as ‘an independent, objective and reliable examination of whether undertakings, systems, operations, programmes, activities or organisations are operating in accordance with the principles of economy, efficiency and effectiveness, and whether there is room for improvement’.

The American Institute of Certified Public Accountants (AICPA) offers the following definition:

‘Performance audits provide objective analysis, findings, and conclusions to assist management and oversight bodies with improving program performance and operations, reducing costs, facilitating decision making and contributing to public accountability’ (<https://www.aicpa.org/content/dam/aicpa/interestareas/business-industryandgovernment/newsandpublications/downloadabledocuments/yellow-bk-audit-jan2020.pdf>).

There is a clear distinction between the performance audit and the financial (compliance) audit. Whilst in the *finance audit* it is a matter of judging ‘whether financial operations have been legally and regularly executed and accounts are reliable’ and the focus is therefore on the financial transaction, the bookkeeping and the control mechanisms, a *performance audit*, with its broader focus, tries to find out whether or not the funds ‘have been used with economy, efficiency and effectiveness’ (European Court of Auditors 2017, p. 10).

Given that, as stated above, the *classical audit* mainly checks the completeness, accuracy and transparency of financial balance sheets, reports etc. and compliance with regulations and laws, this method is highly formalised and geared to a prescribed catalogue of criteria. The audit procedure itself, in turn, must be oriented toward certain framework criteria and standards.⁷

Whilst audits were used for decades to help ensure ‘that public money was spent prudently and wisely’, the *performance audit* has a much wider range of tasks which has been expanded in the last two decades with extra aspects ‘to include policy and regulatory evaluation, areas where conventional auditing expertise is of only limited help’ (Pierre, Peters and de Fine Licht

comparisons between evaluation and controlling can be found in Gerlich (1999, p. 15), in Birnkraut (2011, pp. 8–9), to some extent in Buhl (2015, p. 15) and in Nelke (2021, p. 5).

⁶ The terms ‘audit’ and ‘auditing’ are used synonymously here.

⁷ Cf., for example, the Performance Audit Manual of the European Court of Auditors (<https://www.eca.europa.eu/en/Pages/ecadefault.aspx>) or the American Institute of Certified Public Accountants (AICPA) (<https://www.aicpa.org/>), who make reference to the so-called ‘Yellow Book’, the Generally Accepted Government Auditing Standards (GAGAS) of the United States Government Accountability Office (2021).

2018, p. 726). Whilst in the past auditing was said to be an “‘ex-post” fact process’, ‘ensuring that public servants had acted within their mandate and in accordance with regulatory frameworks’, the new tasks called for more creativity. The activity of auditing is said to have changed ‘from a passive, reactive exercise to more proactive and prescriptive control over actors in the public sector’ (ibid. p. 727), to underpin processes of reform.

Apart from finance and performance audits, there are audits that are deployed in quality management systems (cf. Chapter 4.6) in order to investigate whether or not certain processes, standards and statutory requirements are being complied with, for example to be certified. For this, the product or performance quality of an enterprise can be examined, or an entire system, for example the management system (DIN EN ISO 19011).

A distinction is made between *internal audits*, which do not follow any particular regulations and are conducted on a voluntary basis, and *external audits*, which are conducted for independent certification agencies (pursuant to the specifications of DIN EN ISO 10011) by trained, independent auditors using predefined test items and standardised checklists.

The process of a *performance audit* comprises three essential phases:⁸

The planning of an audit should ensure that the audit is feasible, that it has clear and reasonable aims, that it defines a realistic and solid audit approach, and that the necessary resources are made available.

- (1) The first *planning steps* consist of investigating whether or not the audit project can be carried out at all in the first place. For this, existing feasibility studies are assessed with regard to the planned audit. Whatever happens, an investigation should be made into whether or not the audit can lead to meaningful results within a reasonable period of time. In detail, it is also necessary to clarify what the audit is supposed to accomplish; what the situation is with regard to the availability of the data required for the audit; whether or not the management of the organisation to be assessed already has performance measurement criteria; whether performance measurement criteria are cited in publications that could be used, or whether data for comparison are available in other organisations; and how high the cost of the resources required for data collection is going to be.

Parallel to the feasibility study, there should be a preparatory ‘survey’. This should not be taken to mean the social science term in the sense of a standardised survey, but a kind of *exploratory field trip* by way of interviews and informal talks. The aim is to develop an initial, fundamental understanding with regard to the object of the audit, especially with regard to the organisation’s strategic and operative objectives and their measurability, the organisational structure, existing legal and administrative framework conditions and the internal control structure. In addition, persons outside the organisation that is to undergo the audit should also be consulted, and that also includes those who may hold a critical view of it.

Furthermore, the *aims of the audit* are to be specified more closely and clarified, and the methodological procedure for the actual audit itself developed. Then, the audit plan

⁸ These three phases are to be found generally in all audit institutions, though there are, of course, deviations in the detail. Cf. for example AICPA (<https://www.aicpa.org/>) and European Court of Auditors: Performance Manual 2017 (https://www.eca.europa.eu/Lists/ECADocuments/PERF_AUDIT_MANUAL/PERF_AUDIT_MANUAL_EN.PDF).

should be agreed with the management of the organisation that is going to be assessed. An important objective here is to encourage the willingness of those involved and thus the acceptance of the audit.

At the end of the planning process the objective and the exact course the audit is to follow are finalised in a written document.

- (2) *Field work*: The purpose of the investigation phase is to collect an adequate amount of relevant and reliable *data*, which enable all the statements made in the audit report to be verified. The facts ascertained with the aid of the *investigation procedures* followed – ‘what is’ – are then assessed applying the predefined investigative criteria – ‘what should be’. Thus the *results of the investigation* consist of the comparison of the findings with the standard – expressed in the form of a question or criterion – and their analysis. The results are communicated to the audited body.

The *investigation itself* comprises both an analytical and a communicative aspect. The analytical process relates to the collection, analysis and assessment of data, whilst the communicative process includes the presentation of the results to the audited body and the subsequent exchange of opinions and information.

- (3) *Reporting*: The results of the audit are put together in a report, whereby the organisation that is assessed – audited – is given the opportunity to comment on it. The principal contents of such a ‘*performance audit report*’ are descriptions of the background of the object assessed, objectives of the audit, description of the methodological procedure including the assessment criteria on which it was based, results of the audit, conclusions drawn and recommendations for the management.

SUMMARY

- ✓ Like controlling, the audit developed from accounting.
- ✓ Alongside the classical finance audit, which focuses on questions of the ‘correct’ use of funds, the performance audit is gaining significance more and more, because it uses criteria of effectiveness as well as economic criteria and criteria of efficiency.
- ✓ So-called certification audits are necessary in quality management and ISO certifications; they are carried out in accordance with strictly predefined criteria.

4.3.2 Audit and Evaluation in Comparison

It is not only the use of the instrument evaluation that has been booming for decades now, a fact which Ray Rist and Nicoletta Stame (2011) put nicely in a nutshell by giving their article the title ‘From Studies to Streams’, but also that of auditing, which inspired Michael Power to name his book *The Audit Society* (1999). It is clear that there has not only been a massive increase in certification audits in the context of QM systems and ISO certifications, but also that the governmental ‘supreme auditing institutions’ – as described – are expanding their activities more and more and spreading out to cover new subject areas in order to be able to play an active role in policy change and administrative reforms in the public sector; areas which were, in former times, primarily reserved for evaluation.

Pierre, Peters and de Fine Licht (2018, p. 726f.) believe that this trend is detrimental to evaluation: ‘However, while auditing has increased its focus on performance auditing and in many countries become an agent of administrative reform, conventional evaluation has become marginalised’ (ibid. p. 728). ‘The heyday of policy evaluation’, they say, is over, adding that policy evaluation had been a flop with its time-consuming analyses, simplistic indicator models, coupled with the hubris of being able to say ‘whether a program was working or not, but could also provide answers about how to make it work’ (ibid. p. 728). Instead, they go on to say, auditing is increasingly taking on the role that evaluation used to play, which has led to a situation in which the ‘evaluation function is now being performed by auditing organisations, consultancies, as well as by institutionalised versions of performance management’ (ibid. p. 733). Performance auditing, they add, has gained more and more importance since the 1980s in the context of NPM, whilst evaluation has lost importance. This finally leads the authors to the conclusion: ‘*auditing is to a large extent “the new evaluation”*’ (ibid. p. 736).⁹ However, the authors fail to produce any empirical evidence of this loss of importance on the part of evaluation.

Other studies offer evidence of the exact opposite (cf. Furubo, Rist and Sandahl 2002; Widmer, Beywl and Fabian 2009; Rosenstein 2013, 2015; Jacob, Speer and Furubo 2015; Stockmann and Meyer 2016, 2017a, 2021b, 2021c). Investigations into the institutionalisation of evaluation in the countries within Europe, North and South America reveal an unbroken trend toward more evaluation (cf. Stockmann, Meyer and Taube 2020a, 2020b; Stockmann, Meyer and Szentmarjay 2022). That applies in particular to the political system, in which NPM has proved to be a special driving force.

In spite of the enormous political, social and cultural differences between the countries examined, there are primarily two factors that are responsible for that development: on the one hand, the efforts made toward rational public administration in the context of NPM, and on the other, EU requirements relating to the distribution of subsidies, which are also connected with the ideas of NPM. (Stockmann and Meyer 2021b, p. 416)

One thing certainly not in dispute is that *performance auditing* has undergone massive change and now looks at topics and methods that were formerly reserved for evaluation. Having said that, this is not the case to the same extent in all countries. Whilst *supreme audit offices*, e.g., in Scandinavia, the USA, Great Britain, Switzerland and New Zealand conduct not only performance audits, but also evaluations, the audit offices of other countries, such as the German Federal Court of Audit (BRH) restrict themselves to classical success monitoring. The aim is to investigate the compatibility of administrative practice and organisation with the existing statutory requirements (Müller 2011, p. 29). Political decisions are not assessed by the BRH. Special attention should be paid here to three criteria:

- (1) *Economy*: review of frugality in administrative action.
- (2) *Efficiency*: review of the economic deployment of manpower, material and other resources.

⁹ Emphasis by the author.

- (3) *Effectiveness*: review of the effectiveness of the performance as it relates to the objective, whereby the ‘possibilities that the BRH itself has to investigate such impacts (...) are tightly limited in terms of time and personnel’ (Müller 2011, p. 30).

If we take a closer look at performance audit and evaluation,¹⁰ it becomes clear that the theoretical assumptions behind the individual operations, the specific questions and the methods used for data collection and analysis still diverge markedly.

As long ago as the middle of the 1980s, Eleanor Chelimsky (1985) mapped out the main *differences* between (external) evaluation and audit. As described, these two concepts have come closer together in the last few decades, mainly because the accounting offices have relaxed their audit rules and procedures for concepts of evaluation and social science methods, which used to be very strict, as they now focus more than they used to on the impacts of programmes and measures. Brooks (1997, p. 113) cites the following as a reason for this: ‘The traditional auditing approach has simply proved too limited in its ability to respond to the needs of important clients such as state legislatures’.

It is true that increasingly, evaluators too are using methods which are traditionally deployed in auditing because questions of efficiency and the cost-benefit ratio have gained in importance, but in general it must be said that auditing has moved closer to evaluation than the other way round. There has been a kind of ‘liberalisation’ of the traditional approach. This process has been supported in some countries by the personnel policy of government audit institutions, which have appointed social and administrative scientists and influential evaluators to key positions (cf. Müller 2011, p. 34; Pierre, Peters and de Fine Licht 2018, p. 730; Owen 2021, p. 263). Even if separate ‘cultures’ of auditing and evaluation continue to exist, this process of rapprochement is likely to continue in the future. That means that the differences that still exist could now become more blurred,¹¹ also with regard to evaluators’ and auditors’ *understanding of their roles*:

In dealing with efficiency, auditors may function as if they were management consultants; and when they deal with effectiveness, especially in terms of policy and program matters, they may come close to becoming policy analysts or policy evaluators. (Bastoe 1999, p. 105)

A comparison of *similarities and differences*, which of course can only be rather simplistic here, must first take into account the fact that auditing and evaluation come from completely different traditions (cf. Figure 4.2). As mentioned at the outset, auditing has its roots in 19th-century bookkeeping. Evaluation did not develop in its modern social science manifestation until the 1960s. Yet unlike audit, evaluation was ‘concerned with theory and explanation rather than regularity and compliance’ (Pollitt and Summa 1997, p. 88). Evaluation continues to be more of a science than an administrative procedure. It avails itself of all kinds of social science concepts and methods, turns its attention to all the phases of the political process and has a very wide range of tasks. Auditing, by contrast, is a clearly formulated concept and may have a higher degree of professionalisation, but a far more limited range of tasks.

¹⁰ In that comparison, reference is made to the performance audit, not to evaluations, which are conducted by some supreme audit offices.

¹¹ Cf. the remarks by Brooks (1997, p. 113 ff.) on a ‘blended approach’ of auditing and evaluation. Also Schwandt and Halpern 1988; Davis 1990; Brooks 1996; Leeuw 1996.

Whilst evaluations undertake input, output and outcome observations; investigate unintended and long-term effects; or the sustainability, usefulness and social relevance of programmes and measures and the connections between individual impact areas, performance audits concentrate mainly on the implementation phase, the achievements (outputs) and the direct impacts (outcomes). In audits, this question continues to be at the centre of interest: ‘How does the “condition” under investigation compare to some “criterion” of performance?’ (Bastoe 1999, p. 104; cf. Brandt 2009, p. 101; Müller 2011, p. 30). That is to say that auditing is normative, comparing the target – ‘what should be’ – with the current status – ‘what actually is’. Thus an audit is a retrospective activity.¹² It is true that evaluation too is often designed retrospectively, in particular in impact and sustainability analyses, but it can concentrate just as well on processes that are actually going on, or can even be used prospectively.

Whilst auditing as a rule uses predetermined assessment criteria, the clarification and development of criteria are often explicit parts of the evaluation process. Differences can also be ascertained as regards the recipients of the report results. The BRH, for example, first sends the results of its audits to the audited bodies and then, most importantly, to Parliament (cf. Müller 2011, p. 31; Pierre, Peters and de Fine Licht 2018, p. 735).¹³ In evaluations, as a rule, a variety of addressees are served, including the client, various different stakeholders and a broader public – but they are seldom addressed directly to Parliament. That means that audits tend to focus more on the legislative (body) whilst evaluations focus more on the executive.

One essential difference between auditing and evaluations consists of the fact, ‘that the former usually target institutions while the latter tend to target programs’, which means ‘that auditing generates a different type of information than evaluation and is of different value to policy-makers’ (Pierre, Peters and de Fine Licht 2018, p. 734). The efficiency of organisations in carrying out programmes, however, is an important variable in the success or failure of programmes. That is why Stockmann’s CEval approach (2006, 2014) incorporates an organisational model: to assess the implementation capacity of an organisation (cf. Chapter 2.2.2).

Frequently, a central difference between audit and evaluation is seen in the kind of questions asked: ‘Audit is interested in whether what has been done conforms – a “what” question. Evaluators endeavor to understand what produces certain desired or undesired effects – a “why” question’ (Pollitt and Summa 1997, p. 89).

Rist too (1989) draws attention to this difference. For him, the logic behind the audit lies in the circumstance that normative questions are answered so that a comparison can be drawn between current conditions and defined standards or criteria. Cause-and-effect issues, by contrast, tend to be disregarded. Evaluation is seen as the broader approach, which looks into descriptive, normative and cause-and-effect issues.

For the isolation of causal factors, and in order to be able to exclude alternative explanations, elaborate methodological designs and statistical procedures are often deployed in evaluation (cf. Chapters 7–10). These two approaches mainly have recourse to different data

¹² The BRH traditionally audits *ex post*, but also carries out ‘ongoing’ audits so as to recognise undesired developments as early as possible (cf. Müller 2011, p. 35).

¹³ Whilst the BRH investigation reports are not accessible to the general public, the main results and recommendations of the BRH are sent annually in the so-called ‘remarks’ to the Bundestag (Parliament), the Bundesrat and the Federal Government (Section 97 [1] of the Federal Budget Ordinance [BHO]). The ‘remarks’ are then published as a Bundestag printed paper and thus become generally accessible after all.

sources. Whilst auditors depend to a great extent on documentary material, evaluators use the whole range of social science data collection methods. Yet here too, a process of relaxation can be observed in the performance audit.

An important difference between auditors and evaluators which, by contrast, can hardly be denied, is related to the role that they assume vis-à-vis the subjects of the investigation. Whilst the auditors – if they are employees of an audit office – are endowed with a sovereign assignment¹⁴ and can insist on the handing over of information, data and/or confidential material, evaluators have to rely on the readiness of the evaluatees to cooperate.

By contrast, more distance, a more reserved atmosphere and sometimes even a certain amount of mistrust are, as a rule, prevalent between auditors and those who are being assessed. Evaluators, on the other hand – depending on the evaluation assignment they are carrying out – have a relatively open, at best even trustful working relationship with the evaluatees and other stakeholders. Their judgements and assessments are often built on the advice and opinions of those involved. If participative – or even empowerment – evaluation approaches are used the evaluatees are even more strongly integrated and the role of the evaluator changes from that of an objective observer to that of a moderator or agent. Such a thing as a participative audit, by contrast, does not exist.

Reference is often made to the fact that auditors attract more attention with their reports and results than evaluators because they are endowed with more power due to their status. Occasionally, indeed, procedures have been prescribed according to which those being audited are not merely under obligation to acknowledge the results stated in the audit reports, but also to implement them (cf. Müller 2011, p. 33). Evaluation literature, by contrast, is full of complaints about how decision-makers ignore evaluation results and recommendations (cf. Chapter 6.4). However, the question of whether audit recommendations really do get implemented more often than evaluation recommendations has so far not been investigated. There is some evidence to the effect that evaluation results obtained with the active participation of those involved have a greater chance of being implemented than those which have been realised with only a modest amount of stakeholder involvement and, above all, without the evaluation being accepted. That would tend to speak in favour of greater chances that the evaluation recommendations will be implemented.

To sum up, it can be said that a series of procedures are understood by the term audit / auditing in the public sector, ranging from *traditional finance auditing*, which limits itself primarily to recording inputs and outputs and thus closely resembles controlling, to the *performance audit*, which has a number of evaluation features and identifies with *similar principles*. Using scientific methods, both procedures make efforts to produce objective and precise results which are of use to decision-makers. Both base their implementation on realistic, well thought out and cost-conscious planning. They also aim to pay heed to the rules of fairness in implementation. This conforms to the fundamental standards of evaluation.

Audit and evaluation are instruments of a modern policy that believes in control and is oriented toward criteria of rationality, which make a contribution to improving the dirigibility and public control of government programmes and organisations.

Serious *differences* result above all from the institutional context of audit and evaluation. Government audit institutions have a high degree of independence, coherence and collective

¹⁴ The members of the BRH, indeed, enjoy judicial independence (cf. Müller 2011, p. 20).

Criteria	Auditing	Evaluation
Origin	19th-century accounting	1960s USA in connection with social programmes, social science concept
Focus	Monitoring (regularity and compliance) Administrative orientation	Theory and explanation, Scientific orientation
Mission (i.e. aims)	Improvement in transparency and scrutability of public institutions	Contribution to solving social problems
Phases of political process	implementation, findings	Planning, implementation, impact
Main tasks	Analysis of implementation process, direct performance (output) and effects (outcome)	Analysis of implementation process and of output, outcome, sustainability, relevance, usefulness and other aspects
Research perspective	Retrospective	Prospective, current, retrospective
Research logic	Normative, target-performance comparison	Impact-oriented (including unintended consequences), cause-and-effect logic
Assessment criteria	Precisely defined in advance applying general standards	Less clearly defined, some selection and clarification still done during evaluation process
Audiences	Legislative	Executive, clients, peer groups, stakeholders, public
Research focus	Institutions, organisations	Strategies, programmes, projects, measures
Research designs	Mainly cross-sectional analyses	(Field) experiments, longitudinal, cross-sectional analyses etc.
Methods	Document analysis, interviews	Whole social science spectrum
Role of auditor / evaluator	Independent	Independent, or integrated – moderating, consultant – role in participatory evaluations
Taking account of stakeholders' interests	Unimportant	Very important
Influence of auditor / evaluator	High: some audits are of a binding nature; access to all information	Low: evaluators have to rely on cooperation with valuees
Findings	Entity being audited obliged to adhere to these; effective	Consequences depend on acceptance by evaluated entity

Figure 4.2 *Auditing and evaluation in comparison*

solidarity: ‘They house a coherent, well-defined professional community, buttressed by internationally accepted standards and strong statutory powers’ (Pollitt and Summa 1997, p. 104; Müller 2011, p. 36; Pierre, Peters and de Fine Licht 2018, p. 735). If evaluation units exist in organisations, their internal institutional significance depends to a great extent on their independence, their professional competence and their status, and on the organisational culture being practised. Evaluations conducted by external units, of course, are not, like audit offices, endowed with institutional powers, let alone statutory authority, and they do not form a homogeneous HR pool. Instead, they are in competition with one another for orders, and that can also have a bearing on their independence (Stockmann, Meyer and Schenke 2011; Pleger and Sager 2016).

What distinguishes evaluators or evaluation units is their profound, methodologically imaginative *modus operandi* on its firm theoretical base, creative and original ideas, social competence and professional skill. Whilst auditors from government audit institutions can, to a certain extent, select the objects and topics they wish to investigate themselves (cf. Müller 2011, p. 36), evaluators are largely forced to rely on what is on offer on the assignment market.

Given the similarities and the complementary relationship ascertained between audit and evaluation, *very considerable mutual learning potentials* may be presumed to exist here, and even more use could be made of them. Even if evaluation is not – as described by Pierre, Peters and de Fine Licht (2018) – in decline, it must be realised that the performance audit is increasingly taking on tasks and adopted methods which were still classical domains of evaluation only a few years ago. So if auditing is not to advance to become the ‘new evaluation’, evaluation itself will need to continue to become more professional.

SUMMARY

- ✓ Audit and evaluation are instruments of a modern policy oriented towards criteria of rationality.
- ✓ Both approaches make a contribution to dirigibility and public control.
- ✓ In recent years auditing has moved closer to evaluation in terms of the tasks it tackles and the methods it uses.
- ✓ Serious differences result from the institutional contexts of evaluation and audit. These cannot ultimately be resolved.

4.4 BENCHMARKING¹⁵

4.4.1 General Introduction

As an approach, benchmarking is quite different to the others dealt with above. *Benchmarking enables performance comparisons* to be made, for which the data produced by monitoring, controlling, auditing or evaluation can be used, regardless of whether they are finance, input, output, outcome, impact or other data. Thus benchmarking is referred to less ‘as a technique of measurement than a way of thinking, a disposition towards comparative assessment and learning’ (Kuhlmann and Bogumil 2018, p. 543), which, in public administration, is an important instrument in the context of the NPM concept. Benchmarking is a relatively new management instrument, which was developed in the 1980s and has hardly changed at all since then in conceptional terms (Delers 2018, p. 36). According to its founder Robert Camp (1994, p. 13), benchmarking is ‘the continuous process of measuring products, services and practices against the strongest competitor or the companies which are viewed as leaders in the industry’. The aim is to reveal the differences in respect of the other enterprises and find out the reasons for

¹⁵ There is a wide range of introductory literature on the subject of benchmarking, e.g., Siebert and Kempf 2008; Bruno 2009, p. 261ff.; Kitzinger and Georg 2016; Delers 2018; Sauro 2018; von Bandemer 2019; Kounev, Lange and von Kistowski 2020.

them so as to be able to make improvements. The *aim* of the process is to advance to become the ‘*best of the best*’. Owen (2021, p. 180) defines it thus:

Basically, benchmarking is the search for best practices that can be applied with a view to achieving improved performance. Benchmarking is a systematic and continuous process of measuring and comparing an organisation’s business processes against those of leaders anywhere in the world, to gain information which helps drive continuous improvement.

Performance comparisons have been known in industry for a long time. The application of the core idea can, in fact, be traced back to the introduction of work at the conveyor belt in the automotive industry in 1916. Henry Ford found inspiration in a large Chicago abattoir when he was on a visit there and saw the pigs suspended from hooks sliding on a single overhead rail from one worker to the next. The transposition of that principle to the automotive industry corresponded to the classical process of benchmarking (cf. Siebert and Kempf 2008, p. 10).

A *benchmark* was originally a physical mark on a work bench, for example to enable pipes to be cut to the same length. It is also the name for a point in a known position in land surveying. As used here, a benchmark is a reference point in a measured best performance.

The *core idea* of benchmarking consists of using existing solutions to problems for the improvement of one’s own products or processes. Basically, this is about the search for better alternatives in one’s own enterprise or others. This is not a question of continually reinventing the wheel, but – given that it already exists – of copying it as well as possible and developing it further. For those reasons, benchmarking sometimes has to face the accusation of merely being a substitute for one’s own creativity and ability to innovate. Particularly evil tongues, indeed, place this instrument close to industrial espionage (cf. Siebert and Kempf 2008, p. 23; Delers 2018, p. 32).

The greater the potential for learning from the partner or object of comparison, the more successful benchmarking can be. That makes it clear that great attention must be paid to the choice of a suitable *benchmarking partner*. Various problems can crop up here: (1) the field of vision for potential benchmarking partners is limited, (2) no criteria exist for the assessment of benchmarking partners and (3) there is no database for the selection of benchmarking partners.

Ideally, the selected benchmarking object should be compared with a ‘best in the class’ or a world-class enterprise. The benchmarking object may, for example, be a product, a service, a method or a process. The ‘best’ product or the ‘best’ way of carrying out a process – ‘best practice’ – is elevated to the standard for the comparison and set side by side with one’s own products or processes. Camp (1994) designates the difference between the performance of one’s own enterprise and the benchmark as the ‘*performance gap*’ which is to be bridged.

The measurement of the performance gap is followed by the investigation of the crucial question: how do the others do it better? In other words, by means of what procedure(s) has the reference enterprise made its way to the top of the ladder? The results of this investigation process are then used for improvements in one’s own enterprise, with the aim of taking over that number one position oneself. *Learning from others* is institutionalised in an ongoing programme. Furthermore, that comparison with others is intended to contribute to making a realistic assessment of the competitive strength of one’s own enterprise.

Based on the comparison partners selected, *three forms of benchmarking* can be distinguished:¹⁶ benchmarking within the same company, benchmarking within the same sector and benchmarking between different sectors (cf. Figure 4.3).

Partner for comparison	partner in the same company or group	partner in the same sector	partner in a different sector
Standard of comparison	internal	external	external
Type of BM	internal BM	BM in the same sector	BM between different sectors

Figure 4.3 Types of benchmarking (BM)

In benchmarking within the same company or group, a comparison is made between individual parts of the enterprise. A prerequisite for that, of course, is that there are structures operating in parallel in the enterprise / group (e.g., one consulting department for profit organisations and one for non-profit organisations), so that a meaningful comparison can be made between them.

With the aid of *internal benchmarking*, the aim is to uncover differences in performance within an enterprise and derive potentials for improvement. However, experience shows that the changes made with the intention of improving performance often turn out to be rather small, because the cultural and organisational conditions in a single enterprise do not demonstrate any major differences. In benchmarking, comparison with external partners is thus preferred. This can be competition-related, in other words it can take place within the same sector, or it can be sector-independent.

Advantages		Disadvantages	
✓	Promotes the idea of benchmarking within the company	✓	Mainly suitable for large enterprises which operate internationally and are organised decentrally
✓	POC in company is known	✓	Maximum performance standard in the enterprise itself
✓	Same corporate structure	✓	No positioning possible in comparison with competitors
✓	Open access to information		
✓	Easy to record data		

Figure 4.4 Advantages and disadvantages of benchmarking within the same company

Benchmarking within the same sector is comparison with a partner from the same sector as the one in which one’s own enterprise is active. In other words, one or more competitors are picked out with whom one’s own enterprise is in direct competition. This form of benchmark-

¹⁶ A distinction is sometimes also made between ‘friendly’ benchmarking, in which the enterprises carry out their activities in partnership, and ‘unfriendly’ benchmarking, when information is gathered purposefully against the will or without the knowledge of the enterprise in comparison. In the latter case, the transition to industrial espionage is clearly accurate (cf. Komus 2001, p. 42).

ing is often conducted by independent third parties. Benchmarking within the same sector has the advantage that competitors offer great potential for comparison and the procedure enables one’s position in respect of the competition to be determined directly, but it is, for obvious reasons, difficult to find competitors who are actually prepared to take part in such a comparison. Worries about divulging competition-relevant information to the opposition usually outweigh the hope that one might gain some advantage by this means. Apart from that, in benchmarking within the same sector it is obvious that no one can get better than the best in the sector, and weaknesses typical of the sector cannot be overcome this way either.

Benchmarking cannot develop its full efficacy until it is in its *trans-sectoral* form, in which any enterprise at all can be adduced as a partner for comparison. This creates the chance of finding completely different solutions that are untypical of one’s own sector. The example cited at the beginning of this section, that of how the automotive industry was able to learn from the meat processing industry, makes it clear what kind of potential benchmarking has.¹⁷ In benchmarking between two sectors, of course, it is not products that can be compared, but only processes, procedures and methods. To make the best of the potentials for improvement, however, it is necessary to look very closely indeed at whether or not the findings really can be transposed into one’s own context. Since the organisations taking part in benchmarking are not in a competitive relationship with each other, they are generally more willing to engage in an open exchange of information.

As mentioned at the outset, distinctions can be made between various different *benchmarking objects*, i.e., products and processes. However, strategies can be objects of benchmarking too.

Advantages		Disadvantages	
✓	High degree of acceptance of results because BM object is the same or similar	✓	Difficult to find a partner because degree of willingness to take part tends to be low
✓	Comparison with direct competitors possible, which promises great benefit	✓	Difficult to procure and exchange information
✓	Positioning is possible as compared with competitors	✓	Sector-typical weaknesses cannot be overcome
✓	Identification of partners for comparison is simple		

Figure 4.5 *Advantages and disadvantages of competition-related benchmarking (BM) within the same sector*

A frequently used method for *benchmarking products* (or services) and technical solutions is ‘reverse engineering’. In reverse engineering, one’s own product and a competitor’s product are juxtaposed. Then the two products are stripped down to their individual parts and compared.

¹⁷ A much-quoted example from recent times is the benchmarking project by the photocopier manufacturer Xerox, which found itself in a threatening position in terms of competition at the end of the 1970s. For the process of internal outbound logistics, the mail order company L.L. Bean was chosen as a benchmarking partner.

Advantages		Disadvantages	
✓	Open exchange of information possible	✓	Difficult to select partners who offer a good basis for comparison
✓	Very high learning potential, because innovative methods untypical of the sector are adduced for comparison	✓	Difficult to compare BM objects
✓	Possible to orient oneself to world's best	✓	Difficulties in transferring 'best practices'
		✓	Great amount of time required
✓	Wide range of application	✓	Positioning in competition not possible

Figure 4.6 Advantages and disadvantages of trans-sectoral benchmarking (BM)

Product benchmarking serves above all to bring about cost reductions via the redesign of existing products and via the realisation of identified possibilities for improvement in the development of new products.

Process benchmarking is the form most often used in practice. Here, processes within the same company, within the same sector or between two different sectors are compared. Above all, this is a question of discovering and understanding the backgrounds and causes of the process sequences which are responsible for the results.

Strategic benchmarking serves to examine and improve corporate strategies with the object of gaining long-term advantages in competition. In particular, this is a matter of trying to gain a leading position in terms of market, product or technology.

Typically, benchmarking goes like this:¹⁸

- (1) Selection of the benchmarking object (e.g., product, process, method, strategy etc.).
- (2) Specification of the objective of the benchmarking study (e.g., cost reduction, quality improvement, boost in innovation etc.).
- (3) Identification of the performance factors that are to be measured on the selected benchmarking object (e.g., process time, process quality, process costs; innovation time, innovation quality, innovation costs etc.).
- (4) Identification of benchmarking partner(s) (e.g., the best in the sector or world market leader).
- (5) Recording of performance and performance processes in the partner enterprise.
- (6) Recording of performance and performance processes in one's own enterprise.
- (7) Identification of 'performance gap'.
- (8) Development of programmes and measures to bridge that performance gap.
- (9) Implementation of measures and monitoring of results.
- (10) Integration of the improved mechanisms (processes etc.) in one's own enterprise.

Although benchmarking can be deployed for all kinds of comparison process and in any sector, its *application* is limited mainly to private enterprises (cf. Kitzinger and Georg 2016,

¹⁸ The benchmarking procedure has been articulated in various degrees of detail. Cf., for example, Camp (1994, p. 20f.); Kotler and Bliemel (1999, pp. 406–408); Siebert and Kempf (2008, p. 70ff.); Kitzinger and Georg (2016, p. 12ff.); Delers (2018, p. 41ff.); Owen (2021, p. 181).

p. 3f.; Delers 2018, p. 9). In the public sector, where benchmarking is used in the context of administrative reforms to create incentives for improvement via performance comparisons, the instrument has not been able to win through, although there were considerable expectations associated with the acceptance of performance comparisons in 2009 in Art. 91 of the German Constitution (GG) (cf. Wollmann 2002, p. 33; Wegener 2004, p. 251; Kuhlmann and Bogumil 2018, p. 544; von Bandemer 2019, p. 671).¹⁹ As long ago as 2011, and again in 2015, the BRH criticised its inadequate implementation. Benchmarking did not get beyond symbolic deployment, and in its practical application it has not, so far, been able to develop any impact (cf. von Bandemer 2019, p. 672).

Benchmarking has been able to achieve a much higher status at municipal level, though it has not advanced to become a standard instrument with blanket coverage.²⁰ Mostly, it was only the initial comparison that was important to the municipalities, so that they could determine their position and/or use it for their ‘political marketing’. Hardly any of them were interested in a long-term, regular exchange of data, since the benefit of this form of controlling and benchmarking was not in harmony with the costs incurred. (Cf. Kuhlmann, Bogumil and Wollmann 2004 for a summary.) So it is no surprise that the quality improvement via comparison with other municipalities, which had been hoped for, largely failed to materialise (cf. Wegener 2004, p. 254).

There are a number of other reasons for this: for one thing, the use of benchmarking in the public sector is more complicated than it is in the private sector, because criteria such as the price, the turnover or the market share are lacking as assessment standards for successful action. For another, the measurement of ‘performance differences’ with regard to outcome and impact calls for the application of evaluations. But evaluations are associated with considerable effort in terms of methodology and time and thus also with considerable cost.

There are also implications that there is a danger that highlighting ‘better’ solutions might lead to a general pressure to demonstrate legitimacy, which could trigger off resistance to benchmarking (cf. von Bandemer 2019, p. 669). Doubts about the comparability and transferability of results are often put forward against benchmarking. Indeed, that argument cannot be refuted, but it often turns out to be pretextual.

Furthermore, the benefit of benchmarking, particularly in the public sector, consists not only of the presentation of transferable results, but also of the creation of transparency. The differences disclosed in the rendering of public services facilitate a discourse about alternative solutions and potentials or needs for action, or indeed they need to be disclosed to make it possible in the first place.

¹⁹ Well-known examples of benchmarking among the German federal states are considered to be the studies on educational outcome as part of the Programme for International Student Assessment (PISA), and the comparisons between institutions of higher education made by the Centre for Higher Education (CHE).

²⁰ This is in spite of the intensive support provided by the Bertelsmann Foundation and the Kommunale Gemeinschaftsstelle für Verwaltungsmanagement (KGSt).

SUMMARY

- ✓ The core idea of benchmarking involves using existing solutions to problems to improve one's own products or processes.
- ✓ The principal investigative question is: 'How do the others do it better?'
- ✓ The aim of benchmarking is to enable one to get to the top position oneself.
- ✓ Benchmarking is an analysis and planning instrument with which products, methods, sequences and structures of operational functions can be compared with one another in order to discover potentials for rationalisation or quality and improve performance.
- ✓ Benchmarking can be undertaken on the basis of comparisons made within the same company or outside one's own company with the best competitors in the same sector or with enterprises in other sectors (best practice).
- ✓ At present, benchmarking takes place mainly in the private sector. In the public sector this instrument has not been able to gain acceptance in spite of being anchored in the German Constitution (GG).
- ✓ Since in the public sector there are no clear parameters such as price, turnover or market share by the application of which the success of an enterprise can be measured, benchmarking is more difficult to carry out in the public or NGO sector, but it is deployable if meaningful output, outcome and impact indicators are selected.

4.4.2 Comparison with Other Approaches

Benchmarking is a method for which the data gathered with the aid of the approaches referred to above can be used for *processes of comparison*. Given that benchmarking has so far mainly been deployed by companies, it is mainly quantitative data that are used for the measurement of performance capability in competition. Beyond that, however, qualitative data can also be used (cf. Komus 2001, p. 46). *Causal analyses* – such as are, for example, common in evaluation – do not have any role to play in benchmarking (or, by the way, in controlling, monitoring and mostly in auditing either). Instead, benchmarking limits itself to the observation of processes in order to identify 'best practices' and then imitate them in the best possible way. The question of what causes them is secondary. It is a matter of learning from the best and implementing what has been learned in one's own organisation in the way that brings the greatest possible benefit.

Benchmarking has a certain *propinquity to evaluation*: the establishment of benchmarks, the measurement of performance factors in one's own organisation and the organisation(s) for comparison, and the assessive comparison of differences as a prerequisite for the improvement of one's own processes (cf. the list of steps in benchmarking featured above), all feature similarities with ex-ante or formative evaluations. A *link with monitoring* becomes clear in the further operations, which have to do explicitly with monitoring the impacts caused by the measures implemented – e.g., improved performance, rationalisation effects etc. (cf. Gebel 2006, p. 1; Owen 2021, p. 180f.).

However, other than in evaluations, for example, no theoretical observations are made or assessment criteria derived therefrom, let alone are stakeholders involved. The comparison criteria are already set to a large extent and orient themselves toward the 'best practices' of the organisation(s) with which the comparison is being made. The bandwidth of evaluative

questions is far greater. Yet in spite of that, benchmarking can be included in the portfolio of evaluative instruments. As mentioned at the beginning, the data obtained with various different instruments can be used for benchmarking depending on the objective.

4.5 SYNOPTIC COMPARISON

The approaches presented are concepts influenced predominantly by *business management* – controlling, audit, benchmarking – or *social science* – in some cases audit, monitoring, evaluation – which can be used for various different kinds of task. Not least, however, they all serve to generate control-relevant information for management so that decisions can be made on a rational basis. These are instruments of a modern company management or policy control oriented toward criteria of rationality. That is why they also play an important role in the context of quality management (models) and so-called ‘new’ governance models such as the NPM approach. Before we look at that, we will give a synoptic presentation of the similarities and differences between the genuine business management concepts covered in this chapter and the monitoring and evaluation approaches explained in Chapters 2 and 3.

Strictly, such a comparison is an impossible undertaking, given that there is no single controlling or audit concept; instead, each of these approaches has numerous nuances and areas of emphasis and is subject to a whole series of developments all the time. But if this venture is embarked upon in spite of that, it is in order to impart a *very rough overview*, which enables certain procedures to be categorised with reference to monitoring and evaluation. It should be clear, however, that the possible gain here in terms of systematisation or categorisation can be cancelled out by a loss of information.

As can be seen in Figure 4.7, these concepts, which originally came from business management, have a number of things in common which distinguish all of them quite clearly from the social science approaches.

Monitoring, as a genuine social science concept, comprises both elements of evaluation – for example, the scientific origin, its deployment mainly in the public sector, the breadth of the reporting field, the assessment criteria used, the consideration given to stakeholders’ interests – and business management concepts. Some things it has in common with the latter, for example, are that it also tends to be more of a surveillance instrument and an ongoing internal process, that the assessments are geared to target parameters and thus retrospective, and that they are directed primarily at the management of a company or an organisation.

A detailed look has already been taken at the things *audit* and *evaluation* have in common and the differences between them (cf. Figure 4.2), so there is no need for a repeat of that here.

The greatest differences in Figure 4.7 are to be seen between *evaluation* on the one side and the *business management approaches* on the other. Whilst the latter look more at internal company processes and mostly neglect external ones, it is often exactly the other way round in evaluation. Criteria for the performance ability of organisations that carry out programmes are relatively seldom assessed, whilst a great deal of attention is paid to the intended and unintended effects.

A similar observation can be made regarding the respective assessment criteria: both groups of instruments pay great attention to effectiveness, i.e., the degree of target achievement. That, however, is just about the end of the list of similarities, because the approaches that come from business management are largely based on parameters of efficiency, which, by contrast,

Features		C	A	BM	M	Eva
Scientific origin	- business management	X	X	X		
	- social sciences		(X)		X	X
Main use	- private sector	X	X	X		
	- public sector		X	X	X	X
Tasks	- surveillance / control	X	X	X	X	X
	- explanation					X
Focus of interest	- internal processes	X	X	X	X	
	- external processes				X	X
Methods	- quantitative	X	X	X	X	X
	- qualitative				X	X
Main orientation	- top-down	X	X	X	X	X
	- bottom-up				X	X
Assessment criteria	- predetermined	X	X	X		
	- variable				X	X
Periodicity	- ongoing	X		X	X	
	- sporadic		X			X
Assessment process	- internal	X		X	X	X
	- external		X			X
Area of research	- narrow (e.g., finance / customer-related)	X	X	X		
	- wide (company-related)				X	X
Research logic	- normative (target / actual)	X	X	X	X	
	- cause and effect					X
Research perspective	- prospective					X
	- current				X	X
	- retrospective	X	X	X	X	X
Target audience	- one-dimensional (e.g., management, legislators)	X	X	X	X	
	- multi-dimensional					X
Assessment criteria applied	- effectiveness		X	X	X	X
	- efficiency	X	X	X	(X)	(X)
	- efficacy		(X)		X	X
	- sustainability				X	X
	- social relevance				X	X
Stakeholder interests taken into account?	- yes				X	X
	- no	X	X	X	-	-
Character of results for management	- binding	(X)	X	(X)	-	-
	- non-binding				X	X

Legend:

C = controlling A = audit M = monitoring
 () = not very pronounced BM = benchmarking Eva = evaluation

Figure 4.7 Comparison tableau

are often neglected in evaluations. To make up for that, the bandwidth used in evaluation is far greater; apart from the criteria cited here, it also covers, above all, questions of effectiveness, sustainability and sociopolitical relevance, with the unintended consequences taking on special importance.

When the other features listed in Figure 4.7 are considered, this difference between the instruments that are based on business management and the social science evaluation

approaches is obvious. Whilst the former were developed more for the needs of privately organised, profit-oriented enterprises – controlling, benchmarking – the others are used more in the public sector and for programme control – monitoring, evaluation. Audit and benchmarking somehow assume a more hybrid position. As far as the origin of the principles and areas of their application is concerned, they belong in conceptional-methodological terms more among the management instruments of the private sector. However, audit and benchmarking are also deployed in the public sector. Benchmarking can be carried out with data that can be obtained using all the methods introduced here. Audits are deployed for purposes of certification in enterprises and also for the review and control of public programmes. Accordingly, the individual concepts have varying degrees of affinity to superordinate management models such as quality management or NPM.

To generalise, it can be said that evaluation is usually more open, more participatory, more comprehensive, more scientific (i.e., more analytical with regard to the causes), less normative, broader as regards the field of investigation, the criteria and methods applied – encompassing quantitative and qualitative aspects – and directed at a complex audience. Having said that, it is also less focussed and less binding for those responsible for control than business management instruments and audit.

4.6 USE OF INSTRUMENTS IN QUALITY MANAGEMENT AND FOR PROGRAMME CONTROL

Unlike the instruments introduced so far, which can be used as management tools in enterprises, administration or politics for rational decision-making, *TQM*, *ISO certifications* and *NPM* are superordinate concepts for ensuring the quality of the products made or services rendered or, generally speaking, the outputs generated.

Quality management models, which declare quality to be the most important determinant in the success of an enterprise, distinguish themselves by strict *customer orientation*, because quality is calculated above all from the assessment of the benefit by the customers. That is why all the processes and structures are geared to the fulfilment of the customers' needs.

Various different models are available for the configuration of quality management. They can be based on a series of standards (DIN EN ISO 9000–9004) or on the central ideas of TQM, which in turn are based on well-known quality awards such as the Malcolm Baldrige Award and the European Quality Award.

What all these models have in common is the systematic reprocessing of the organisational structure, the recording of the processes and the assurance of the quality of the product to be produced or the service to be rendered. With the aid of standards and norms, the aim is to create a uniform information base with which enterprises can be assessed and compared. The decisive assessment criterion is the *quality* of the product or service (cf. Raidl 2001, p. 53; Brügemann and Bremer 2020, p. 1). It may be true that all QM models claim to be deployable in all sectors and branches, but the differences between the producing and the service-providing sectors already create difficulties in their implementation: in the services sector, the immateriality of the service renders a contemplation of their quality more difficult, repeated measurements are impeded by a lack of storability, the consumer is integrated in the

process of service provision so that the result of the performance and thus also its quality are not solely in the sphere of influence of the provider, and there is no possibility to exchange or return goods (cf. Löffler 2019, p. 362; Haller and Wissing 2020, p. 458). The use of QM models in the non-profit sector, which we will be looking at later on, is even more difficult.

*ISO 9000*²¹ is not an instrument for ascertaining a quality level; it defines minimum requirements relating to a quality management system and aims to document the work procedures in their entirety. It is not a quality improvement system, but a quality assurance system. The starting point is the idea that the best possible quality will be achieved if the construction of the product is logically planned, standardised and transparent for all those involved. The ISO standards govern all the areas which need to be defined and organised in a quality management system. Following inspection by an officially certified organisation (e.g., TÜV, DEKRA), compliance with the ISO standards can be certified. The aim of this is to make it clear that the work done in an enterprise is true to quality. Having said that, ISO 9000 certification does not offer any guarantee that all the employees actually apply the rules, that modern quality management is actually practised, or that the defined procedures are actually being optimised.²²

An alternative quality concept is the *TQM approach*,²³ which is not based on fixed standards, but gears all the company processes to the primacy of quality. The management introduces TQM and guides the continuous improvement of all corporate processes, but all the employees are held responsible for quality assurance and development (cf. the comprehensive overview in Stockmann 2006, p. 32ff.; Haller and Wissing 2020, p. 457f.). According to the European version of the model of the *European Foundation For Quality Management (EFQM)*, distinctions are made between five enabler criteria and four results criteria, with each group being divided up into further sub-criteria (32 in all). Applying the criteria for the *enablers* – management, policy and strategy, employees, resources, processes – procedures, activities and processes in an enterprise and the degree to which they are applied can be examined and assessed. The enabler criteria deal with what the organisation does. Given that, when all is said and done, the aim of all quality endeavours is to improve the operating results, the main points systemically assessed on the *results side* are the (operating) results, employee and customer satisfaction and the image outside the company (social responsibility / image). The results criteria deal with the results an organisation achieves by its performance.²⁴

²¹ The whole of the ‘standards family’ ISO 9000ff. has been revised several times in recent years. This set of rules, introduced by the International Organization for Standardization (ISO), aims to render a uniform understanding of QM systems possible. For further information go to www.iso.org.

²² Cf. Hartz and Meisel 2011, p. 62ff.; Zollondz 2011, p. 301ff.; DIN ISO 9000 to 9004 with regard to Europe 2013; Seghezzi, Fahrni and Friedli 2013, p. 193ff.; Hensen 2016, p. 118ff.; Kohl and Neumann 2020, pp. 27–30; Zink, Schmauder and Bengler 2021, p. 196f.

²³ It is very difficult to get a clear overview of the literature on the subject of quality management. Here are a few important titles to choose from: Cristian and Costel 2011; Zollondz 2011, p. 301ff.; Zink 2012; Bovermann 2013; Oess 2013; Schiersmann, Thiel and Pfizenmaier 2013; Seghezzi, Fahrni and Friedli 2013, p. 193ff.; Wilken 2013; Rothlauf 2014; Schmitt and Pfeifer 2015; Kiran 2016; Gómez, Martínez Costa and Martínez Lorente 2017, pp. 88–103; Ross 2017; Schawel and Billing 2018; Gnahn and Quilling 2019; Zedler 2020; Pfeifer and Schmitt 2021.

²⁴ See in detail: <http://www.efqm.org>.

The instruments introduced so far, such as controlling, auditing, benchmarking, monitoring and evaluation are independent concepts which can be used for company or policy management (also largely independent of one another). They do not have to be integrated in quality management systems, but they can be integrated to procure and assess data (information). *Audits* are used very often. Every ISO certification²⁵ is preceded by an audit, which investigates whether or not the prescribed standards are being adhered to. KPI-based systems such as *controlling* can be deployed subsequently for a continuous norm control. Whilst audits, which can lead to certification, are carried out externally – by auditors trained for the purpose – audits in the context of quality management models such as EFQM tend to be organised internally.

Benchmarking also makes an important contribution in the context of QM models by enabling a comparison to be made with organisations that are exemplary in the quality domain and providing incentives for quality improvements. Particularly in the self-assessment undertaken in accordance with the specifications of the EFQM, the ‘objective outside view’ is an explicit prerequisite (cf. Zink 2004, p. 291ff.; Siebert and Kempf 2008, p. 25; Thul et al. 2015, p. 269f.; Hensen 2016, p. 124ff.).

Since the middle / end of the 1990s, interest in QM in public administration and throughout the *non-profit sector* has increased. Attempts are often made to transpose the models developed for the profit sector (such as ISO and EFQM), but they are not always successful. In German public administration, QM is used above all in the social and healthcare sector and the education sector. There are no backed up data on the prevalence of QM in the public sector (cf. Löffler 2019, p. 360). The fact that adaptation may not be that easy to achieve is, on the one hand, to do with the instruments themselves:

A principal difficulty is that the skills and techniques of the private sector are not directly transferable to the public sector. In particular, private sector techniques, with their emphasis on profitability as a measure of effectiveness, are of little value in evaluating policy outcomes (Smith 1996, p. 169).

Moreover, the primacy of German administrative culture, i.e., the legitimacy and expediency of administrative action, which competes with the first principle of any QM concept – customer satisfaction – proves to be a barrier to the implementation of QM in the public sector. Other hurdles are said to be a lack of interest on the part of top-level administrators and policy-makers, a lack of management skills, a lack of motivation to initiate processes of change, forces of inertia in traditional administrative cultures, too few resources etc. (cf. Löffler 2019, p. 368).

Viewed structurally, the fact that QM has so far seen little use throughout the non-profit sector (including public administration) is associated with a number of organisational and situational differences when compared with the profit (private) sector. Figure 4.8 gives a brief summary of the main ones. (For a more detailed explanation cf. Stockmann 2006, p. 44ff.)

In many European countries, there was increasing dissatisfaction with the traditional modes of operation and the performance capability of state administration at the end of the 1980s.

²⁵ If it is a certificate issued in accordance with a set requirement profile (for example, certified in accordance with DIN ISO 9001), the term ‘certification’ is used. But if the certifier of the organisation and the issuer of the stamp of quality are one and the same (as is, for example, the case in the certification of study courses in tertiary education), it is referred to as ‘accreditation’ (cf. Haller and Wissing 2020, p. 468).

Profit (private) sector	Non-profit (public) sector
<ul style="list-style-type: none"> ✓ Aims to make a profit, acts for own benefit ✓ Financed via prices obtained for products and/or services ✓ Clear main goal (e.g. maximisation of profit) ✓ Clear parameters (e.g. profit, dividends, shareholder value) ✓ Open competition ✓ Free choice of providers ✓ Products and services ✓ One-dimensional producer-customer relationship 	<ul style="list-style-type: none"> ✓ No intention to make a profit, benefit for others ✓ Financed via collective contributions (e.g. taxes, payments, donations) ✓ Complex systems of goals ✓ Parameters need to be determined (variable, multi-dimensional) ✓ Competition often restricted, sometimes none at all ✓ Choice of provider often restricted, in some cases no choice ✓ Services ✓ Multi-dimensional level relationships (incl. funders, clients, producers, addressees)

Figure 4.8 Differences between profit and non-profit sectors

Inspired by the principles of a market-oriented economy and oriented toward the new theory of institutions and the public choice theories, which were becoming more and more popular in the world of science, a management model referred to as *NPM* came into being, which saw itself as an alternative to the traditional models ‘in which the administration is defined as a system geared to legitimation, integrated in the policy process (policy model), or as a system geared to the application of rules (bureaucracy model)’ (Schröter 2019, p. 121). Whilst in some countries (e.g., Great Britain), *NPM* was used to propagate a general retreat of the state in favour of private-sector solutions, the ‘new management model’ limited itself to the modernisation of the internal structures of public administration by improving internal management mechanisms (cf. Jann 2019, p. 129). The central aim of this more technocratic approach was to replace administrative management based on inputs (dispensation of resources) by a management oriented more toward achievements (outputs) in order to improve performance and the process of rendering services in public administration. For that, *four strategic aims* are aspired to:

- (1) orientation toward the customer,
- (2) orientation toward competition,
- (3) orientation toward performance and impact,
- (4) orientation toward quality.

The establishment of competition-like structures, orientation toward the requirements and needs of the clients, for whom certain services are rendered, and the creation of quality consciousness makes *performance- and impact-related management* necessary. This is accompanied by a number of methodological difficulties, which cannot be surmounted with the control

and finance instruments usually deployed in administration.²⁶ Comprehensive *controlling and benchmarking processes* based on KPIs have therefore been developed, though they have hardly been able to gain acceptance in the public sector at all (cf. Chapter 4.4). Although they were originally associated with the hope of being able to increase the efficiency and effectiveness of publicly rendered services (output), it would still not have been possible to make any statements about the outputs achieved (outcome and impact) that constitute the true purpose of rendering services, even if the application had been more broadly based (cf. Brüggemeier 2004, p. 375; Löffler 2019, p. 363).

For those reasons, the deployment of evaluations has been in demand for ‘appropriate public sector measurement’ for a long time now (Smith 1996, p. 169). However, so far public administrations have been making heavy weather of recording not only outputs, but also impacts. And the same goes for the *performance audits* carried out in NPM.

Even in Great Britain, where ‘performance measurement’ is considerably further developed than it is in Germany, the focus is more on operational indicators and less on the effectiveness of political measures, let alone their effectiveness (cf. Wegener 2004, p. 259; Kuhlmann 2010, p. 335; Barkowsky 2014, p. 211).

Increasingly, this restriction to business management parameters is criticised as ‘*economic reductionism*’ (Kuhlmann, Bogumil and Wollmann 2004, p. 12; Schröter 2019, p. 117). The mere use of business-related indicators and benchmarking is obviously not sufficient for performance- and impact-oriented governance such as is, for example, necessary in NPM. Looked at internationally, there are some signs that governance is gearing itself more and more to outcomes, but in Germany there is no similar development to be seen (cf. Wollmann 2004, p. 41, 2016, p. 31; Barkowsky 2014, p. 211; Poller et al. 2015, p. 131; Friedländer 2019, p. 79).

To sum up, it can be seen as a *central deficit* that the previously practised indicator systems, performance measurements and benchmarking have up to now been oriented too little toward impacts and hardly say anything about the causes of ‘best practices’. (For summaries of this cf. Kuhlmann, Bogumil and Wollmann 2004; Kuhlmann and Wollmann 2013).

This is where the *great strength* of the evaluation approach is to be seen; its main task lies in the discovery and measurement of impacts and in cause-and-effect analysis. Although ‘new management models’ urgently need information about the performance level attained – the output and impacts achieved – in order to render output or even outcome-oriented management possible at all, the potential of evaluation in the public sector is often not exhausted.

It may well be true that the efficiency of organisations can be improved by controlling and the determination of their position by means of benchmarking, but the impact analysis and the investigation of *cause-and-effect relationships* fall by the wayside. That means that essential potentials for innovation and learning are wasted. Evaluation could remedy this situation. With evaluations, above all, provided that they are designed participatively and trusted by those

²⁶ The NPM approach is not merely confronted by problems of implementation, but also by a wide range of fundamental criticisms, for example to the effect that the administration is being reduced to the status of a service enterprise, the reform perspective is focused on the efficiency of individual companies, the functional interweaving of politics and administration is being disregarded, and a technocratic understanding of problems and reforms is now prevailing which does not do justice to the political-democratic requirements relating to modernisation etc. (cf. Schröter 2019, p. 121f.)

involved, comprehensive assessments of output and outcome can be made and major learning effects achieved. Thus it would seem advisable to deploy both instruments based on business management and evaluations for corporate or organisational management.

Whilst the former serve, above all, purposes of surveillance, in particular that of internal enterprise-related processes, evaluations make it possible to assess external processes of change and analyse cause-and-effect relationships. A *combination* of KPI-based tools (controlling), approaches for continuous performance recoding with quantitative and qualitative data (monitoring and evaluations), which not only assess target achievement but also scrutinise the targets themselves, comparing intended and unintended effects, as well as employing further criteria (such as sustainability and societal-political relevance) for assessment, and exploring the causes of observed changes, could be the key to successful *organisational management*, particularly in the public sector, and could lead to a more comprehensive utilisation of innovation and learning potentials than before.

SUMMARY

- ✓ Business-management-based concepts such as controlling, benchmarking and audit are often used in quality management and company management in the private sector.
- ✓ In both the state sector and the non-profit sector overall, the deployment status of private-enterprise management instruments is still experimental.
- ✓ Evaluation tends to be deployed less often as an instrument of control / management in the private sector, but more often in the governmental and non-profit sectors.
- ✓ Looked at overall, the control instruments that have been developed are not being used adequately and considerable amounts of innovation and learning potential are being wasted.
- ✓ With a combination of KPI-related instruments, approaches for recording performance on an ongoing basis, and evaluation, the management of programmes and organisations could be improved considerably.
- ✓ Above all, because of the way it investigates impacts and causal connections, evaluation lends itself especially to the support of target- and impact-oriented management.