Benchmarking labour market performance and labour market policies: theoretical foundations and applications

Schütz, Holger; Speckesser, Stefan; Schmid, Günther

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
SSG Sozialwissenschaften, USB Köln

Empfohlene Zitierung / Suggested Citation:
Schütz, H., Speckesser, S., & Schmid, G. (1998). *Benchmarking labour market performance and labour market policies: theoretical foundations and applications.* (Discussion Papers / Wissenschaftszentrum Berlin für Sozialforschung, Forschungsschwerpunkt Arbeitsmarkt und Beschäftigung, Abteilung Arbeitsmarktpolitik und Beschäftigung, 98-205). Berlin: Wissenschaftszentrum Berlin für Sozialforschung gGmbH. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-128821

discussion paper

FS I 98 - 205

Benchmarking Labour Market Performance and Labour Market Policies: Theoretical Foundations and Applications

Holger Schütz, Stefan Speckesser,
Günther Schmid

Juni 1998
ISSN Nr. 1011-9523

e-mail:
holly@medea.wz-berlin.de
specki@medea.wz-berlin.de
gues@medea.wz-berlin.de

ZITIERWEISE / CITATION

Holger Schütz, Stefan Speckesser, Günther Schmid

Benchmarking Labour Market Performance and Labour Market Policies: Theoretical Foundations and Applications

Discussion Paper FS I 98 - 205
Wissenschaftszentrum Berlin für Sozialforschung 1998

Forschungsschwerpunkt:
Arbeitsmarkt und Beschäftigung

Research Area:
Labour Market and Employment

Abteilung:
Arbeitsmarktpolitik und Beschäftigung

Research Unit:
Labour Market Policy and Employment

Wissenschaftszentrum Berlin für Sozialforschung
Reichpietschufer 50
D-10785 Berlin
e-mail: wzb@wz-berlin.de
Internet: http://www.wz-berlin.de

Summary

Over the last few years, "benchmarking" advanced to a key word in organisational development and change management. Originally, benchmarking was a tool in business studies summarising the process of comparing your own with a similar organisational unit (mostly the competitor) in order to improve the competitive position. Benchmarking must be distinguished from purely analytical methods of comparison: First, performance indicators must be developed which differ from traditional design. Secondly, an organisational unit must be found which can be classified as the "best performer" concerning the chosen indicators. The comparison then aims at finding options for the improvement of your own organisational unit in different hierarchical levels with the objective to progress in the position of the "best performer".

It is understandable to investigate whether or not benchmarking procedures are also applicable in non-profit and public organisational units. In this field, one of the principal research areas for new comparable analytical tools and alternative performance measurement can be seen in the international comparison of labour market policies and their performance outcomes especially after the Amsterdam treaty on international harmonised employment policies: Mutual learning turns out to be inevitable in the European situation which is characterised by converging employment policies as well as by nationally diverging capacities to cope with the labour market problems.

This discussion paper tests the possibilities of applying benchmarking procedures to the international comparison of labour market policies (LMP) and policy outcomes. We conclude that benchmarking can be performed and should be tested in the monitoring of labour market policies. We indicate both the chances as well as the risks of benchmarking: On the one hand, the introduction of benchmarking could lead to the development and improvement of processes and indicators in LMP evaluation. On the other hand, we clearly point out the danger occurring if the strategies of the "best performers" are used as a blueprints in different institutional contexts.

Following the analysis of benchmarking, we also open a methodological discussion about the performance of employment systems. We then demonstrate alternative performance indicators for policy outcomes. We suggest a procedure enabling the simultaneous measurement of even contrasting or conflicting policy goals. As appropriate example, we perform a routine of processing indicators for different policy goals in one single measure with the aggregation to a surface measure.

Zusammenfassung

"Benchmarking" avancierte in den vergangenen Jahren zu einem Schlüsselbegriff in der Organisationsentwicklung und im Management. Es benennt in seiner disziplinären Heimat, der Betriebswirtschaftslehre, den Vergleich der eigenen mit einer anderen Organisationseinheit zum Ziel der Verbesserung der eigenen Wettbewerbsposition. Benchmarking unterschiedet sich daher von rein analytischen Vergleichen auf zwei Arten: Erstens werden zunächst Leistungsindikatoren entwickelt, die sich von traditionellen Qualitätsindikatoren unterscheiden. Zweitens wird eine vergleichbare Organisationseinheit gewählt, die bezüglich der gewählten Indikatoren als "best-praktizierende" eingestuft wird. Der Vergleich, der über verschiedene Hierarchieebenen verfolgt wird, soll den Raum für Verbesserungen der eigenen Organisationseinheit aufzeigen, um selbst in die Position des "Bestpraktizierenden" aufzurücken.

Es ist naheliegend, dieses analytische Instrumentarium ebenso auf nicht privat-wirtschaftliche Organisationseinheiten anzuwenden. Der internationale Vergleich von Arbeitsmarktpolitik als instrumenteller und Arbeitsmarktperformanz als abhängiger Variable entwickelt sich in Folge der Beschlüsse von Amsterdam dabei zu einem Kernbereich neuer analytischer Instrumente und alternativer Indikatoren der Leistungs-bewertung: Vor allem in einem Europa mit unterschiedlichen Geschwindigkeiten bei der Lösung von Arbeitsmarktproblemen und der gleichzeitigen Konvergenz der nationalen Arbeitsmarktpolitiken wird dieser Vergleich zum Zwecke gegenseitigen Lernens notwendig.

Das vorliegende Diskussionspapier prüft die Anwendbarkeit von Benchmarking im internationalen Vergleich von Arbeitsmarktpolitik. Es kommt dabei zu dem Schluß, daß Benchmarking angewandt werden kann und sich im Monitoring der Arbeitsmarktpolitik bewähren sollte. Es verweist dabei sowohl auf die Chancen, die in der Weiterentwicklung methodischer Bewertungsverfahren liegen, als auch auf die Risiken des Benchmarking bezüglich der Schwierigkeit, in unterschiedlichen institutionellen Kontexten "beste Praktiken" wie Blaupausen anzuwenden.

In der Folge der Auswertung des Benchmarking-Ansatzes eröffnet das Papier auch eine methodische Diskussion über Leistungsfähigkeit von Beschäftigungssystemen und entwickelt alternative Indikatoren der Messung von Politikergebnissen. Es schlägt dabei ein integratives Verfahren vor, konfligierende Ziele von Arbeitsmarktpolitik gleichzeitig zu messen und zu bewerten. Die Aufbereitung von verschiedenen Maßzahlen der Arbeitsmarktperformanz zu einem komplexen Maß wird exemplarisch anhand eines Flächenmaßes verdeutlicht.

Contents

List of Figures

List of Tables

Benchmarking Labour Market Performance and Labour Market Policies: Theoretical Foundations and Applications[1]

A new buzzword has recently entered the debates on labour market policies: benchmarking. At the European level in particular, there is growing concern with and interest in 'benchmarking'. For instance, the European Commission dedicated the whole of its 1997 Joint Employment Report to this subject. This increasing interest gives rise to a number of questions. What is the essential meaning of benchmarking? How is it defined? What kind of methods and tools are applied in benchmarking? Are these distinct from other comparative approaches? What does benchmarking mean in the field of labour market performance and policies? Have benchmarking exercises already been undertaken in this area? What can it offer and where do its limits lie?

This paper sets out to address these questions and is organized as follows. The first part outlines and discusses the background to benchmarking, together with related definitions and methodological issues; it also includes a section on benchmarking in the public sector in general. This introduction serves as a necessary foundation for the second part of the paper, three sections of which will be devoted to the benchmarking of labour market performance and labour market policies. The benchmarking of labour market performance is the subject of section 2, which discusses methodological approaches and introduces the employment systems approach as a theoretical framework. In the third section we turn to the benchmarking of labour market policies and suggest the transitional labour market approach as a complementary theoretical framework. The potential and limits of the benchmarking of labour market policies will be debated. In section 4 we turn more specifically to benchmarking within the framework of the European Employment Strategy (until recently Essen Strategy) and investigate the fields of long-term unemployment, youth employment, equal opportunities between men and women and job creation. Our main focus here will be on measurement issues and the development of suitable key indicators for benchmarking. In section 5 we introduce the so-called 'radar chart'-approach, which we consider to be a useful tool not only for the benchmarking of the aforementioned policies, but also for evaluating the performance of public employment services; several examples will illustrate the range of applications. We conclude the paper (section 6) with a summary of the main points and some remarks on the future prospects for benchmarking in employment policies.

---

[1] This paper integrates and extends two (of three) studies that the WZB has contributed to the joint study 'Benchmarking Employment Performance and Labour Market Policies' by the Employment Observatory RESEARCH Network (Tronti 1998) DG V (Directorate-General for Employment, Industrial Relations and Social Affairs) of the European Commission has supported the study financially. The Members of the RESEARCH Network are: Prof. B. Gazier (SET-METIS, Paris), Prof. D. Anxo (CELMS, Gothenburg), Prof. L. Tronti (Fondazione Brondolini, Roma), Dr. J. Philpott (Employment Policy Institute, London), Prof. S. Houseman (W.E. Upjohn Institute for Employment Research, Kalamazoo, USA), Prof. M.Nitta (Institute of Social Science, Tokyo) and Prof. G. Schmid (WZB).

1. Benchmarking in the Private and Public Sectors: Concepts and Methods

Benchmarking was originally developed in private-sector management and, according to a very popular working definition (Camp 1989: 12), 'is the search for best practices that lead to superior performance'. The following more formal definition devised by the Rank Xerox Company is also often quoted: 'The continuous process of measuring our products, services, and business practices against the toughest competitors or those companies recognized as industry leaders' (ibd.: 10). A good definition applicable to both the private and public sectors reads is given by Cowper and Samuels (1997: 11): 'Benchmarking as an efficiency tool is based on the principle of measuring the performance of one organisation against a standard, whether absolute or relative to other organisations'. There are many other similar definitions. Rather than discuss them all, we will begin by considering the various types of benchmarking. In this section we will concentrate firstly on the business-oriented literature and then on benchmarking in the public sector.

1.1 Types of Benchmarking

Just a cursory glance at the management literature makes it clear that that a distinction has to be made between the different types of benchmarking, although it is equally clear that these various types are not defined in the same way throughout the literature. Nevertheless, the distinction between internal, external and functional benchmarking as the basic forms seems to be fairly uncontroversial, with 'generic' benchmarking sometimes included in this category (cf. Camp 1989: 60-65; Karlöf and Ostblöm 1994: 62-67; Watson 1993: 105-209).

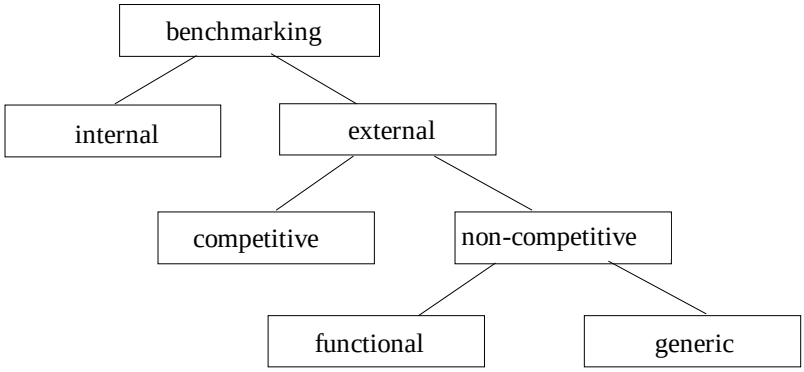
Internal benchmarking takes place between units or sub-units of the same organization; it may be used as an instrument of explicit internal competition, but was not necessarily designed as such. *External benchmarking* seeks to compare a firm with organizations with the same or very similar characteristics. The benchmarking partner may be a direct competitor or a firm operating in other market segments. The main characteristic of external benchmarking is the high degree of comparability between the organizations or products being benchmarked. *Functional benchmarking* denotes essentially the comparative analysis of specific tasks, functions or processes independently of product, sector, branch or market segment. For Karlöf and Ostblöm (1994) functional benchmarking is synonymous with best-practice benchmarking, i.e. finding the best performer in a particular field. Functional benchmarking in dissimilar industries may be advantageous because of higher acceptance (fewer problems with confidentiality, greater objectivity) (Camp 1989). It presupposes some kind of logical comparability which is determined, as Camp argues, by specific product characteristics, but not the product or the industry itself[2]. If product characteristics or type of industry

[2] One example would be issues around product handling, in which the weight, size, shape and fragility of products will determine the usefulness of comparisons.

do not play any role at all in comparability[3] the purest form of benchmarking, *generic benchmarking*, can be used (ibid.). In fact, the difference between functional and generic benchmarking is quite small (see also Karlöf and Ostblöm 1994).

A slightly different view is offered by Watson (1993: 108f; 147, 165), who considers functional benchmarking to be limited to sectoral activities; generic benchmarking is the approach to follow if neither sectoral nor competition boundaries restrict applicability. Additionally, both functional and generic benchmarking are understood as non-competitive concepts since they were not intended to provide competitive advantages over rivals. Watson's conceptualization of benchmarking involves the data source on the one hand (internal/external) and the relationship between the benchmarking partners on the other hand (see Figure 1).

Figure 1: Basic Types of Benchmarking



Source: derived from Watson 1993

One example is the now very famous and much-discussed use of generic benchmarking by Xerox (see Camp 1989; Watson 1993: chap. 8; Bogan and English 1994 26f; Leibfried and McNair 1996: 144f). Having realized in the 1970s that the company was falling drastically behind its competitors, the large, technologically-driven Xerox Corporation became the first company to introduce generic benchmarking (among other things). The initial benchmarking partner L.L. Bean, a much smaller retailer and mail-order warehouse company, had not much in common with Xerox at first glance, but served Xerox very successfully as a superior model in warehouse operations and logistics. Xerox then gradually expanded across-the-border benchmarking; in the late 80s, about 230 performance areas were benchmarked outside the company's own industry. The Xerox Corporation's 'resurrection' from downright underachievement in the late 70s to excellence in the late 80s, when the company won the Malcolm Bridge National Quality Award, is indeed most remarkable.[4]

---

[3] A good example is the management of wage/salary accounts, which is comparable across all sectors and industries.

[4] The Xerox example clearly indicates that comparability should not be rejected too hastily. On the contrary, identifying generic processes that are carried out better in different types of organisations seems to be a challenge. Moreover, the high degree of institutionalization and embeddedness of the benchmarking approach and methods at all levels and units of Xerox is also worthy of note. The

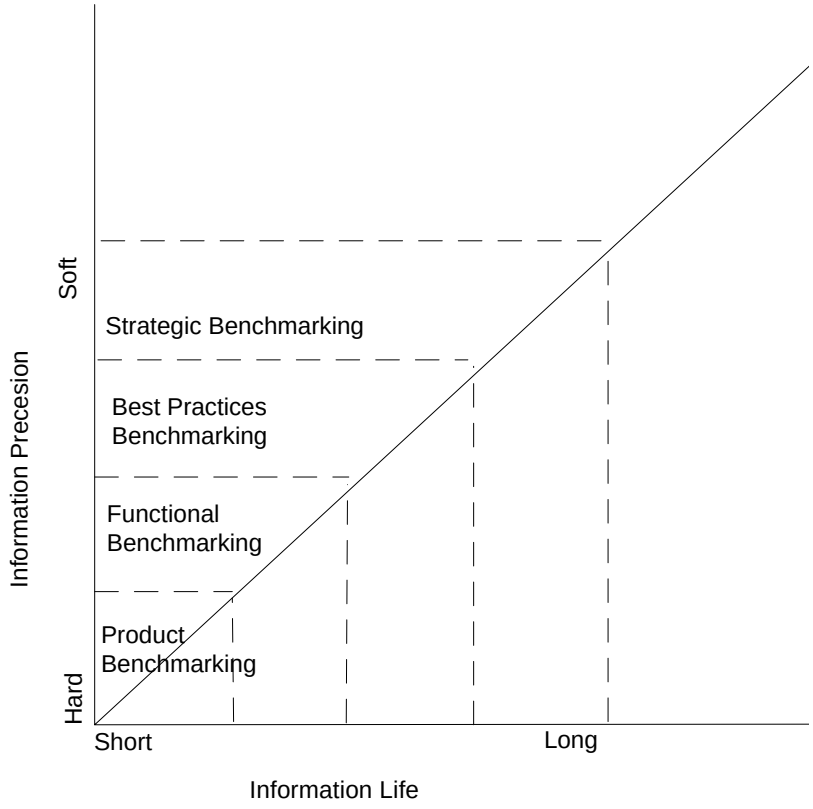
An alternative typology of benchmarking is offered by Miller et al. (1992), who distinguish four types of benchmarking which they label product, functional, best-practice and strategic. In their view, the most traditional type is product benchmarking, that is the practice of carefully examining or "tearing down" a product and comparing it with a benchmark product from a direct competitor. Their definition of functional benchmarking is much the same as the one outlined above: the processes of the best performer are used as a comparative guideline in order to produce continuous improvement. Best-practice benchmarking extends functional benchmarking to the level of management and its implementation strategies. Finally, according to these authors, strategic benchmarking[5] focuses more on the essential organizational aims and objectives to be achieved. Hence it is argued that organisations should start with strategic benchmarking, in order to provide "a context and a rationale that enhances the effectiveness of the other three types" (ibid.: 25).

From our point of view, it does not appear very useful to confine (the definition of) best-practice benchmarking to management practices. Notwithstanding this limitation of their definition, however, the analysis provided by Miller et al. gives a good insight into the relationship between different purposes, time horizons and qualitative details in these types of benchmarking. For instance, strategic benchmarking offers long-lasting but not very precise information since the focus here is on long-term goals. In contrast, product or functional benchmarking procedures should lead to quite precise information about actual processes that may indicate a need for rapid intervention (see Figure 2 below).

involvement of employees is very reminiscent of TQM approaches.

[5] The management literature on strategic planning and strategic management places a great deal of emphasis on interaction with the relevant environments in order to achieve the essential long-term goals of the organization ('mission', 'vision') (see, for instance, Bryson 1993; Naschold 1995: 103-107). Strategic benchmarking has to be related to this context.

Figure 2: Precision versus Life of Benchmarking Information



Source: Miller et al 1992: 26

To sum up provisionally, there are in principle no limits to benchmarking in the private sector since it can be applied to inputs, outputs (products), methods, processes, short-term and long-term goals. The most important prerequisite is to guarantee adequate comparability corresponding to the goals of the benchmarking process. Definitions and typologies of benchmarking differ partly and to some extent, at least, they will be all suited to their specific purposes. For our purposes, however, we tentatively suggest results ('product') and process ('functional') benchmarking as prototypes capable of being adapted to specific research designs and questions.

1.2 Tools and Methods

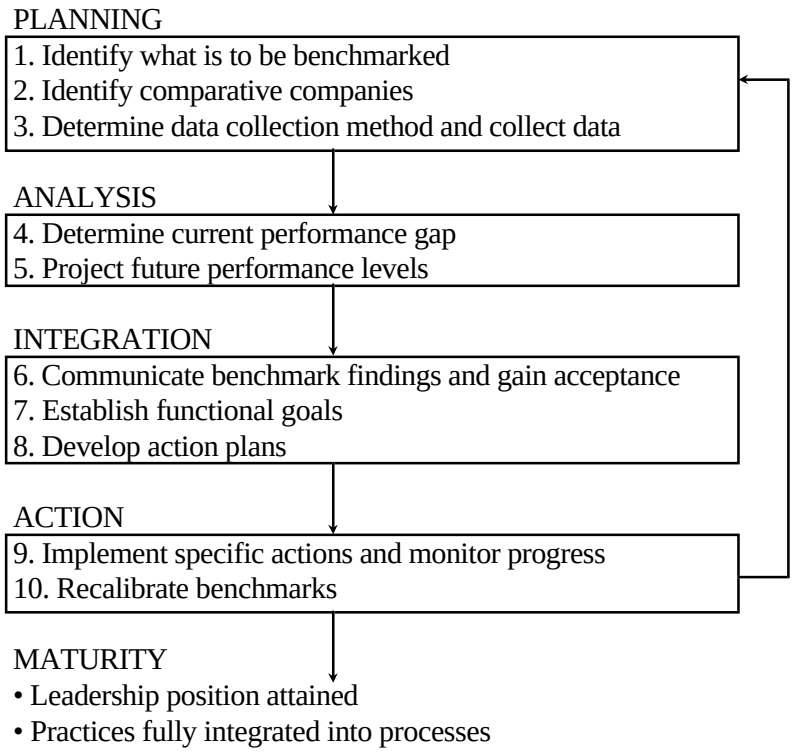
In this section, we will outline the benchmarking process, consider some measurement issues and discuss the factors critical to the success of benchmarking.

1.2.1 The Benchmarking Process

There is a high degree of consensus in the business literature as to the various steps in (a complete) benchmarking process (see e.g. the contributions in Rolstadås 1995). The

common- sense benchmarking cycle, which originates in the early benchmarking activities of the Xerox Corporation, is displayed in Figure 3[6].

Figure 3: Steps in the Benchmarking Process



Source: Camp 1989

The 'feedback loop' from step 10 to step 1 is essential for the institutionalization of benchmarking, which should in turn facilitate organizational learning. Adherence to the main procedural logic is also important for success.

1.2.2 Benchmarking and Performance Measurement

Benchmarking in private companies is explicitly connected with alternative performance measurements. The first condition for the comparison and benchmarking of organisational units is to recognise them as complex systems in which many complicated processes are taking place simultaneously. This is what makes it necessary to apply complex indicators and alternative measurement procedures, as studies of benchmarking in the private sector strongly underline (Bogan and English 1994: 44 ff.). It follows that

[6] It should be noted that some authors prefer to mention only five essential steps (for instance, Karlöf and Östblom 1994): 1) determination of what is to be benchmarked 2) identification of benchmarking partners 3) data generation 4) data analysis 5) implementation into results. Other authors use seven or eight steps (e.g. Boxwell 1994). Notwithstanding the number of and slight variations in the steps involved, the logical structure of benchmarking remains the same.

innovative methodological options and alternative measurements should also be developed and implemented in measuring excellence in public-sector organizations and employment systems.

In business studies, the traditional performance indicators are cash flow, cost, sales, liabilities, debts and profitability. Such indicators will be less useful for process improvement as they usually provide a snapshot of performance at a given time but do not show what happens over time and which processes can be optimised. It has to be realized that indicators of this type will not always provide information on the decisive factors 'behind the scenes' or on how processes are interlinked with others (Bogan and English 1994: 45 f.).

In management research, therefore, the need for more complex performance indicators that include "interaction terms" has been discussed since the early eighties (see Peters and Waterman 1982: 64 ff). Furthermore, it is understood that performance measurement necessarily follows normative guidelines. In this sense, effective benchmark indicators should reflect the most important operational aspects of a business process, function or system, with a mix of proactive and reactive indicators being preferable[7] (cf. Bogan and English 1993: 46-47). The identification and understanding of these 'critical success factors' (i.e. the components that have the most impact on success or failure) is central to the entire benchmarking endeavour (see also Watson 1993; Karlöf and Ostblöm 1994). In sum, benchmarking needs broad performance indicators that reflect the organization's overall performance and opens up perspectives for further structural analysis. Methods of measurement that reduce complex performance concepts to specific sub-tasks and questions are preferable from a benchmarking point of view (cf. Watson 1993: 59 ff; Bogan and English 1994: 49).

This objective is reflected in the vast number of specific methods and models that have been devised for the widest possible range of applications. Any attempt to give just a superficial overview of these diverse methods would be futile (see only Rolstadås 1995). However, some of the benchmarking tools in widespread use should be mentioned: There is for for instance the so-called Z-chart, which illustrates current performance gaps and future performance levels (cf. Camp 1989; Bogan and English 1994: 101-105). Process mapping to depict detailed work flows (e.g. Leibfried/McNair 1996: 57-59) and two-dimensional mapping in scatter plots of performance capabilities or criteria (cf. Boxwell 1994: 60-61; Watson 1993: 33) are also popular. Another interesting measurement tool was developed by the Kodak Company. Kodak, itself a good example of successful internal benchmarking, introduced an innovative comprehensive performance indicator illustrating the company's entire operation and the gap between actual and desired benchmark performance ('Measure Matrix Chart', M^2) (for details see Bogan and English 1994: 58-61). This instrument is very attractive because of its wide range of possible applications.[8]

---

[7] Proactive or preventive indicators are forward-looking, foreshadowing a system's future outcomes. Reactive or lagging performance indicators provide information on the completed performance of a system over a given time period.

[8] This Matrix Measure was also an inspiration for our own work; see section 5 of this paper.

Another very different, though widespread method of performance management is the use of quality awards (see also section 1.3.2). The prototypical quality awards in the private sector are the Malcolm Baldridge National Quality Award (MBNQA) in the United States, the Deming Prize and the European Quality Award (EQA) (hereafter: Löffler 1995). Furthermore, since the MBNQA explicitly awards absolute excellence (best performance) in quality management and achievement, it would also seem to be the prototypical benchmarking award (and is frequently mentioned in the benchmarking literature). Although all three awards focus on quality improvements, there are important conceptual differences. The Deming Prize is not really competitive and is based on the principle of recognition (of a defined relative performance), whereas the MBNQA is definitely competitive (only six possible winners per year) and requires certain absolute performance levels to be achieved. The European Quality Award is based on a "middle-of-the-road" concept that uses both the recognition and competition principles. In terms of assessment criteria, the Deming Prize places more emphasis on process improvements using 'hard' data. In contrast, the Baldridge Award stresses customer satisfaction and is thus mainly based on 'soft' data. The EQA again takes a middle road. Since organizations can considerably enhance their reputations by winning quality awards, they can be regarded as a very powerful benchmarking concept.

1.2.3 Crucial Features of Benchmarking

What is crucial for benchmarking success? One fundamental conclusion is prevalent throughout the literature, namely that *the most important thing is to understand correctly the central goal of benchmarking, which is to learn and improve.* This sounds simple, but in fact addresses the complex issue of how organizational learning is best generated. In benchmarking, learning occurs through the implementation of positive changes based on the identified causes of performance gaps. In other words, the benchmarking cycle needs both to be completed and to be permanently on-going.[9] In the private sector, benchmarking is clearly used not only as an analytical tool but also to provide the impetus for translating new insights into action.

According to survey results collected from successful benchmarking companies, the factors that are critical to the success of benchmarking are (Cross and Iqbal 1995):

- strong commitment on part of senior management ;
- strong emphasis on thorough and complete preparation;
- a high degree of discipline and adherence to the elements of tasks;
- permanent institutionalization of benchmarking, which is not to be conducted solely for problem-solving;
- integration of benchmarking into company priorities and planning;
- availability of adequate time, staff and finance (resource-intensiveness of benchmarking).

---

[9] Organizational learning is of course one of the subjects in interdisciplinary social science research that has attracted most attention in the 90s. For insights from a benchmarking point of view see in particular Bogan and English 1994; Watson 1993.

These findings are widely confirmed in other publications. Rigour in all phases of the process is considered decisive, but the planning phase is frequently identified as the most important. Other features of successful benchmarking emphasised in the literature are team-based approaches and education and training for the benchmarking team. One fundamental mistake is to confuse learning with copycatting (trying to adopt instead of adapt). Furthermore, resistance to benchmarking often has to be overcome. One widely mentioned difficulty is the so-called 'not-invented-here-syndrome', i.e. the rejection of insights originating outside the company. Commitment to benchmarking is most likely to be disseminated throughout the whole organization if employees are involved in the process.

1.3. Benchmarking in the Public Sector

Benchmarking in the public sector is a fairly recent development encouraged by the almost universal trend towards modernization and reform strategies based on notions such as 'New Public Management'. Public-sector reform based on the concept of New Public Management (NPM) began in the early 80s and is now widespread in many OECD countries (for empirical evidence see OECD 1993ff; Naschold 1996). Though NPM resists precise definition because of the diversity of theoretical inputs (mainly public choice theories and management science and literature) that reflect the variety of empirical models and sub-tasks, the main thrust of NPM can be summarized as follows. Firstly, the challenge of NPM is to increase the legitimacy, effectiveness, efficiency and economy (called the 'three E's' in the British debate) and quality of public-sector organizations. Second, NPM seeks to redefine the scope and tasks of public-sector activities (in a more limited sense: the public-private mix of service provision). In this context, benchmarking belongs primarily - as in the private sector- to the field of quality policies. Secondly, it is supposed to be a suitable means of introducing competition into the public sector (Löffler 1996: 145).[10]

1.3.1 Benchmarking and Quality Policies in the Public Sector

This section will offer a brief overview of quality policies before turning to more specific issues in public-sector benchmarking. The three main approaches to quality policies in the public sector are generally said to be quality control, quality assurance and total quality management (TQM) (Oakland 1993; Bovaird 1996). There are also three different approaches to quality management, namely the quality accreditation, inspection and competition approaches (ibd.: 48f). These approaches are summarized in Figure 4.

[10] But if benchmarking is to be used as a means of introducing competition, then it must logically be confined to public-sector organizations with the same or similar goals, products and characteristics (Löffler 1996).

Figure 4: Quality Approaches and Systems in the Public Sector

Approach	Contents
Quality control	Maintenance of product, process or service quality through monitoring and detection and elimination of causes of quality problems.
Quality assurance	Prevention of quality problems through planning and systematic management (systems).
Total quality management (TQM)	Comprehensive approach to ensuring continuous improvement in structures, processes and outcomes, in which all parts and members of the organization are to participate and take responsibility for meeting customers' wishes and needs.
Quality accreditation approach	Assigned to the ISO 9000 series which provides several standards organizations have to comply with
Quality inspection approach	An external regulatory body usually responsible for accrediting and licensing an organisation's operations.
Quality competition approach	Rankings and benchmarking concepts using acknowledged standards for comparison; e.g. quality awards

In empirical terms, these approaches are seldom, if ever, found in their "pure" forms but rather in several hybrid forms. From a benchmarking point of view, neither pure quality control, quality accreditation nor quality inspection are adequate options, since they lack any element of continuous improvement or organizational learning. The rationale of these approaches is somehow inconsistent with demands for greater flexibility and responsiveness in public service agencies towards clients' and customers' needs, since they rely heavily on formal criteria. Moreover, external auditing (quality inspection) might be seen as a threat in the organization concerned if its members are not adequately involved in issues relating to performance improvement (Bovaird 1996: 49). However, the quality accreditation approach may possibly contribute to performance improvements, as the latest ISO-Standards (ISO 9004) are regarded as more flexible and useful for public sector organizations (see Löffler 1996: 148-150, for a brief discussion). Nevertheless, it is TQM[11] and the quality competition approach that lie at the heart of benchmarking. Judging by the number of quality award schemes now in existence (many of them inspired by private-sector examples such as the Malcolm Bridge Award), quality competition seems to be a fairly relevant form of public-sector benchmarking. These schemes will be surveyed briefly in the following section.

1.3.2 Quality Awards in the Public Sector

By the mid-nineties, 16 relevant public-sector quality award schemes had been institutionalized in 12 of the OECD countries[12] (Löffler 1995: 36-39). The majority of

---

[11] For a discussion of the relation between TQM and benchmarking see Swift et al. (1995). Prospects for and limitations of TQM in the public sector are discussed at length by Oppen (1995); see also Löffler (1995: Chapter 3); Morgan and Murgatroyd (1994).

[12] One-off events, very small-scale or still evolving schemes are not included here. International awards, such as the Bertelsmann Prize, are also excluded. Some of these awards are for public and

these schemes seek to identify excellence in public-sector organizations on the basis of a number of specific criteria such as service, problem solving, forward-looking management, performance awareness, quality management and standards. A second objective of these awards is to give public sector organizations a tool for self-audit and assessment in order to enhance quality awareness and performance. Quality awards may thus be a helpful tool for creating a more positive attitude towards organizational learning. Like similar private-sector schemes, these public-sector awards often combine 'objective' with 'subjective' performance criteria. All the award schemes considered here have an NPM focus, although there are some striking differences. For instance, in the *Italian Prize for Innovations in Relations between the Public Administration and the Public*, transparency, innovativeness and awareness of citizens' needs are stressed and judged by public employees. In contrast, the *British Charter Mark Award* Scheme emphasizes market values ("value for money"), while efficiency and awareness of customers' needs are ideally to be assessed by customers and taxpayers. In the USA and Canada, customer satisfaction is given even greater emphasis and is the predominant quality criterion. In Germany, the *Speyer Award* involves public employees and customers in the assessment of quality processes and the results achieved by public-sector organizations. In other words, these quality award schemes or benchmarking approaches all reflect different concepts and understandings of public accountability which in turn suggest that NPM can follow a number of different paths (cf. also Wirth 1991). Before turning to labour market issues in the next part of the paper, some of the interesting approaches to public-sector benchmarking not yet considered will be examined briefly.

1.3.3 Other Approaches to Public-Sector Benchmarking

Cowper and Samuels (1997) perceive three approaches in use in the public sector, which they term standards, results and process benchmarking (see also Trosa 1997). The last two correspond to product and functional benchmarking in the private sector, whereas standard benchmarking simply denotes the setting of a standard of performance which an effective organization could be expected to achieve.

In Britain, one example of standard benchmarking is the internal self-assessment carried out by means of a rating questionnaire in Next Steps agencies against standards laid down in the private-sector UK Quality Awards model. A second example is the *Citizens' Charter* which aims to raise performance standards and awareness of customers' needs in public-sector units and agencies (cf. Cowper and Samuels 1997; for a critical view see Pollitt 1994; Oppen 1995). The Charter standards can be understood as benchmarks for clients and customers, since they provide an easy yardstick by which to measure actual staff performance; they are also benchmarks for employees which cannot be easily undercut.

In Sweden, rating methods were adopted in the early 1990s as a means of introducing continuous improvement into the financial management of local government

---

private-sector organizations. Nine of those 12 awards are based on the competition principle and 2 on the recognition principle, while one is a hybrid type.

agencies (Dahlberg and Isaksson 1997). This interesting example of process benchmarking combines top-down control from central government with bottom-up self-assessment by local agencies. Since the rating method is a fairly recent innovation, outcome evaluations (from a benchmarking point of view) have yet to be produced. Moreover, an annual benchmarking system for local finances, based on comparison of performance indicators, is already in place. Improving the comparability of data and the comparison process itself is still the core task in this benchmarking activity. Rating methods are also used to benchmark the Swedish Budget Process and fiscal performance. In sum, attempts are obviously being made in Sweden to establish benchmarking as an open-ended, continuous improvement process.

2. Benchmarking Labour-Market Performance

Benchmarking should lead to two results (Watson 1993: 35): (1) concrete performance measures (indicators) suitable for the purpose of comparison; (2) identification of the success factors ('process catalysts') crucial to achievement of the desired performance. And, most importantly in the context of this latter outcome (see section 1.2.3), the essential task of benchmarking is to promote organizational learning and improvements. If benchmarking were confined to the construction of performance indicators, it would simply be part of a monitoring process. However, this is obviously not the primary purpose of benchmarking. Thus it would be senseless to use the term 'benchmarking' merely for monitoring activities. The fundamental objective of benchmarking is to identify success factors and facilitate organizational learning. This basic aim needs to be adapted to the context of labour market performance and policies. So, firstly, how is benchmarking to be conceived with respect to labour market performance? We start our answer by addressing the relationship between benchmarking and existing methodologies for measuring labour market performance (section 2.1). On this basis, we propose the comparative analytical approach to employment systems as a suitable theoretical framework for benchmarking labour market performance; the most important implication of this choice is the need to construct relational benchmark indicators (section 2.2).

2.1. Labour Market Performance : The Methodologies of Measurement

The assessment of labour market performance involves not only the measurement and analysis of facts and figures but also the *assessment* of performance. In this sense, benchmarking is not different from *evaluation*, and therefore includes all the available and feasible methods in the evaluation toolbox.

In empirical evaluation research, a broad distinction can be made between interpretative-descriptive and scaled-causal methodologies (Lassnigg et al. 1994: 3-4). In employment and labour market policy evaluation, a wide range of either the one or the other type of methodology is used, but combinations of the two are also widespread. The last point is particularly true for studies that seek to obtain an 'entire picture' of national employment policies or labour market performance. For theory-driven *target-oriented evaluation research* that includes process analysis (policy formation, implementation, take-up), impact studies and some kind of cost-benefit-analysis, a combination of

qualitative and quantitative methods will be essential (Schmid, Schömann, O'Reilly 1996a). In contrast, traditional programme evaluation concentrates on the outputs and outcomes of single programme policy target variables and therefore needs less methodological variety. However, in order to assess the *overall labour market performance* (including the institutional efficiency) of regional, national or transnational economies, it will not be sufficient to rely on a single or very small set of methodologies. Thus, the benchmarking of European labour market policy will necessarily require a combination of causal inference based on various statistical methods and plausible inference based on case study evidence and interpretation.

The following methodological approaches common in labour market policy evaluation can be distinguished:[13]

- *Micro-level impact analysis in an experimental or non-experimental design* measures or estimates the outcomes of programmes on individuals relative to the intended targets and to the situation that would (presumably) have prevailed if the programme had not been put in place.
- *Aggregate impact analysis* measures the outcome of programmes on aggregate indicators such as wages, employment or unemployment at regional or national level.
- *Survey research*, particularly that based on panel studies, is an important research instrument for longitudinal (flow) analysis; it complements process-generated data on take-up gathered by programme administrators by directly seeking the views of 'clients'; it is also useful, therefore, for detecting unintended programme effects.
- *Policy formation and implementation analysis* is essential for linking measured programme outcomes to specific attributes of the programme structure and implementation. It can identify the underlying or explicit policy targets, which are often ambiguous or even contradictory, provide information on different policy choices and may detect policy formation and implementation failures which affect policy effectiveness and efficiency. This strand of analysis has no determined methodology. Thus it may include historical and contemporary narrative description, expert interviews, content analysis of written documents, secondary analysis, network analysis, and even formal models such as game theory and bargaining theory.
- *Cost benefit analysis* (CBA) compares the total benefits with the total costs of programmes and provides a basis for alternative policy choices.

All of these methodologies can be applied in a *comparative framework*. Depending on the subject and purpose of the evaluation, these methodologies will be applied in different specifications (e.g. econometric models, cross-sectional or longitudinal comparisons etc.) and combinations. Thus there is no methodological blueprint which fits all cases. Nonetheless, under 'ideal-type' conditions a target-oriented evaluation of national labour market performance would include: (1) the problem diagnosis, (2) an analysis of

---

[13] For a comprehensive view of the 'state of the art' cf. the contributions in Schmid, O'Reilly and Schömann 1996, and Schmid et al. 1997.

contextual factors that affect the problem structure as well as policy choices, (3) monitoring and process evaluation, (4) impact analysis, (5) a balancing of costs and benefits through CBA (including sensitivity tests) (see also Schmid, Schömann and Schütz 1997).

As far as the introduction of benchmarking into this framework is concerned, it can be argued, firstly, that impact studies and CBA are necessary conditions of 'really good' benchmarks. Only such studies can guarantee that the 'benchmarks' are controlled for unintended and negative side-effects such as deadweight, substitution, displacement, crowding-out or selection effects. Additionally, allocational and distributional effects could be estimated and assessed by CBAs, though CBA methods are, to some extent, controversial. Secondly, good and fairly comprehensive monitoring systems are a necessary precondition for benchmarking. This is all the more important in public services and policies, where processes, organisational efficiency and outputs are far less easily comparable than in the private sector. Thirdly, the chosen benchmarks must take into account national differences in problem and implementation structure and corresponding variations in implementing, for example, the European Employment Strategy.

To sum up briefly, our main argument in this section is that the assessment of labour market performance should encompass a comprehensive evaluation strategy of which *benchmarking* should be an integral part. In other words, the quality of the benchmarks will depend largely on the quality of pre-existing evaluation research. Furthermore, since benchmarking was originally designed as a '*why-tool*' rather than a '*what-tool*' or a '*how-tool*', it clearly makes sense to rely on a wide range of data and methods for the methodological foundation of benchmarks. In fact and in practice, however, a full range of evaluation studies and techniques will never be available for all aspects of labour market performance. Consequently, more pragmatic studies that make use of, for instance, cross-national ranking procedures, will retain their justification and importance (c.f., for example, 'classic' studies with a remarkably broad conceptual base such as Therborn 1985; Schmidt 1986; Rowthorn and Glyn 1990). It should be borne in mind, however, that ranking activity should not be confused with benchmarking, regardless of how highly developed the approach might be; a ranking exercise can reveal the best (or worst) performers, but cannot fully identify reveal the crucial factors in determining success and failure. We now turn to a brief outline of the employment systems approach, which we consider to be an adequate theoretical framework for the complex task of benchmarking.

2.2 The Employment Systems Approach

Employment systems are understood here as the set of institutions and policies that simultaneously determine the level of unemployment and of employment.[14] These institutions act as filters, suggesting certain reactions to external shocks or challenges and more or less excluding other, theoretically possible ones.[15] In turn, employment

[14] For a more detailed exposition cf. Schmid 1998.

[15] On the significance of institutions cf., among others, Garrett and Lange 1995; North 1991; Schmid

systems are characterised by the interaction of two subsystems: the production system and the labour market system.

It is in the *production system* that decisions on production are taken. These decisions depend on interest and exchange rates, technological innovations, effective demand and the cost of production factors. Changes in these parameters are determined by actors whose decisions are, in turn, made within a framework of institutionalised rules: by central banks, (Schumpeterian) entrepreneurs, private households, treasury officials, bodies representing various interests and executive authorities. The rules of the game concern, among other things, the autonomy of central banks, the taxation of primary income and profits, the regulation of market entries, the research & development infrastructure, and so on. From this point of view, unemployment or a poor employment performance can be seen as the result of unrealised or uncompetitive production caused by an unfavourable institutional framework for the production system.

Viewed from this angle, part of the long-term rising trend in unemployment in all industrialised countries can be readily explained. In the 1970s and 80s, growth rates fell by half in virtually all industrialised countries, with the exception of Japan. Industries with the highest productivity increase are no longer those in which employment is expanding, as was the case in the 1950s and 1960s. On the contrary: in many industries in which employment levels have hitherto been high, the price elasticity of demand is declining because saturation points have been reached. In consequence, investment to serve larger markets is not worthwhile and large numbers of jobs are lost. International price competition worsens the situation, and there is not yet any sign of a new long-term (Kondratieff Cycle) economic cycle in which jobs will be created over a sustained period through the application of new information and communications technologies.[16]

In addition to the general decline of growth in industrialized countries, *Europe has a qualitative growth problem*. European production systems seem to be less innovative than the American one.[17] There is much evidence to suggest that the regulation of product markets slows down reactions to the opening-up of new market segments. There are also signs that the most recent winners in terms of growth rates in real GDP (for instance the US, UK, Australia, Canada) have profited from heavy investments in information technology as well as from deregulation or proper reregulation in this sector (Motohashi and Nezu 1997). Moreover, there are also indications that monetary and financial policy is not sufficiently well co-ordinated. Money deposits still earn higher returns than real investments, and labour is taxed too highly relative to consumption and wealth.[18]

This is not the place to investigate this aspect of employment systems in any greater detail. However, these brief remarks were necessary because there is a current tendency to focus solely on the labour market in the search for a scapegoat for the desperate

---

et al. 1992; Schmid 1994.

[16] Cf. Appelbaum and Schettkat 1995; for a somewhat more optimistic view, see Freeman and Soete 1994.

[17] Cf, for example, OECD 1994; Freeman and Soete 1994; Lehner 1996.

[18] Cf., among others, Dornbusch 1994; Commission 1993.

employment situation in Europe. All studies of labour market performance have to keep in mind this interrelationship between production systems and labour market systems, to which we now turn.

As the comparison of growth in the US and Europe indicates, decisions on production are not necessarily followed by decisions on employment. These decisions are made in the labour market. The rules and incentives that lead to employment decisions constitute what we term the *labour market system*. From this perspective, unemployment can be seen as the result of unrealised or misplaced employment. At least four institutions play a role in such decisions, and all of them interact with each other:

- Firstly, the private *household system*, which offers alternatives to paid employment, shapes cultural attitudes towards work and places constraints on the volume of time available for paid work.
- Secondly, the *industrial relations system*, in which the conflicting interests of the various labour market actors come up against each other and whose rules and power relationships determine, in particular, the level and structure of wages.
- Thirdly, the *education system*, which produces general knowledge, learning skills and vocational qualifications and determines the limits of occupational mobility and flexibility.
- Fourthly, the *social security system*, which affects employment decisions in a variety of different ways: (1) through the state as employer, which can offer alternative employment in social spheres outside the market; (2) through the largely state-regulated benefit system, which offers alternatives to earned income for workers in certain risk situations (unemployment, sickness, old age); (3) through the regulation of certain aspects of the employment relationship, such as dismissal protection, fixed-term contracts and working-time arrangements.

Labour market policy can influence employment decisions through all four of these institutional channels: by providing systematic information and advice (job placement); by subsidising wage costs; by eliminating skills shortages; by creating publicly-funded "bridges" into employment or altering the level and duration of benefit payments and, finally, by deregulating or re-regulating employment relationships.

Thus employment systems are very complex institutional arrangements. Their very complexity rules out unidimensional theories of unemployment; equally, however, they cannot be regarded as an arbitrary conglomeration of possible institutional factors. They usually form a coherent functional framework, in other words, an employment policy configuration or *employment regime*, that has developed over time and has regional and national characteristics.

One such configuration, for example, is often referred to as *competitive capitalism*, which denotes the predominance of market mechanisms in decisions on production and employment, as exemplified by the USA. Another is known as *coordinated capitalism*, which refers to the close co-ordination between state and business in decisions on production and employment that characterises Japan, for example. A third configuration is commonly referred to as *welfare capitalism*, which alludes to the important role played

by social security systems in decisions on production and employment in most European countries. The competition that used to exist between capitalism and socialism has now been replaced by competition between these variants of capitalism. There are considerable differences in this respect within the European Union. Not only are these differences more visible because of the competition between the various regimes, but they are also relevant to the question of whether a new, independent and successful European model will emerge from this competition. Will that model develop more in the direction, for example, of the Netherlands or more in those of Germany or the United Kingdom?

Benchmarking can be of great help in finding answers to this question. The employment systems approach, however, has one important implication for proper benchmarking in contrast to conventional benchmarking: indicators have to reflect relationships between and within the various employment systems. Single indicators are misleading and, therefore, 'forbidden' since conventional aggregate performance indicators do not reveal functional equivalents of policies or trade-offs between multiple labour market goals. Typical examples of relational indicators are *employment elasticities* or *unit labour costs*. One way to elucidate relationships within complex aggregate indicators is to use shift-share techniques to break the indicators down into their separate components. For instance, the disaggregation of employment growth into the components of growth rates in the working-age population, labour force participation and unemployment gives important insights into the structural, behavioural and political components of these indicators. A powerful example for the demand side is the decomposition of GDP growth into growth rates of productivity and volume of work in terms of hours and employees.[19] This distinction (of structural, behavioural and political factors) is of special importance for the (comparative) analysis of the labour market performance of different employment systems and will be essential for benchmarking in the future.

[19] For an application see for instance Houseman 1995, Schmid 1997, Werner 1997.

3. Benchmarking Labour Market Policies

The goal of this part is fourfold. Firstly, in connection with and as a specific supplement to the previous section on employment systems, we recommend the concept of transitional labour markets as an explicit normative-analytical framework for the benchmarking of labour market policies. Secondly, despite the fact that we have already argued in favour of the benchmarking of labour market performance and policies, we discuss briefly the applicability and feasibility of benchmarking approaches in labour market policies and point out some of the general and specific problems associated with such approaches. Thirdly, we offer a brief overview of the few empirical examples to date of benchmarking of labour market policies. Finally, from the material and arguments presented so far, we draw an interim conclusion that seeks to define a concept of benchmarking that takes account of both the theoretical objectives and the empirical findings.

3.1 The Concept of Transitional Labour Markets

The transitional labour market approach (Schmid 1995, 1998, 1998a) is a more dynamic concept of labour market performance than conventional approaches. The most important premises underlying this approach are, first, the assumption that a return to the high growth rates and levels of employment of the 'golden age of capitalism' is not possible and, for ecological reasons, not even desirable. Second, the dramatic changes in the age composition of the population in mature industrial societies require a new social contract, especially in relation to the financing of social security. Third, normative aspirations of gender equity and equality call for a new gender contract which replaces the male breadwinner model with a partnership model.

For these and other reasons, the traditional definition of full employment is obsolete. The concept of transitional labour markets offers a new definition which assumes optional breaks in the working career as a 'normal' condition of contemporary and future working life. Such breaks are to a certain extent even a desirable policy target since they would help to (re)distribute the available job opportunities more evenly. Average working time over the life cycle would be reduced substantially[20], but actual working time would vary considerably according to economic conditions and phase of the life cycle. Institutionalised transitions between various working times and/or between different employment statuses would also enhance new forms of labour market flexibility that would enable individual or household labour market options to be reconciled with other activities such as education, further training, social work or cultural and political involvement.

---

[20] From a level of 38 to 30 hours a week or 1800 to 1300 hours a year as a medium-term perspective; at the moment, only the Netherlands comes near to this 'goal'; the OECD Employment Outlook 1997 (Table G, p. 179) reports an average (contractual, not actual) working time of 1372 hours for those in dependent employment.

Because of these normative goals, the most general performance criteria in the transitional labour market approach are the existence and configuration of the institutional bridges that facilitate transitions between (and within) employment, unemployment, education or training and private household activities; the transition to retirement would be gradual and not abrupt. The concept of transitional labour markets also has a tremendous impact on the method used to measure the 'routine' performance of labour markets, with the emphasis shifting from *dynamic indicators* to stock indicators. The following examples are an illustration of this strategic change in the measurement of labour market performance.

3.1.1 The Transition Matrix

Within the framework of transitional labour markets, the incentive effects of different institutional arrangements (for instance unemployment insurance systems, active labour market policy regimes) on labour market actors and their impact on the functioning of labour markets can best be examined in terms of a labour market transition model. Flows in 16 possible directions can be distinguished:

From \ To	Employment	Unemployment	Policy Measure	Inactivity
Employment	1	2	3	4
Unemployment	5	6	7	8
Policy Measure	9	10	11	12
Inactivity	13	14	15	16

Apart from serving as a guide for statistical labour market monitoring systems, the transition matrix can help to identify interesting performance indicators, such as the transition rate from unemployment to policy measures (cell 7) for a certain category of unemployed, for instance those unemployed for more than six months. Other interesting performance indicators for labour market policy would be, of course, the transition rates from participation in a policy measure into employment, unemployment, another policy measure or inactivity after the measure ends (cells 9 to 12), as well as the transition behaviour one year into the measure. Finally, transitions between different employment statuses in a period (say a year) provide interesting information, part of which could be used for benchmarking. Distinctions can be drawn between, for instance, waged and salaried employment and self-employment, full-time and part-time employment, or between permanent and fixed-term jobs.[21]

3.1.2 Balance Sheets of Inflows and Outflows

Stocks of employment or unemployment are always the result of inflows and outflows to and from one specific status. Balance sheets of inflows and outflows, and especially the dynamic of those balance sheets over a time series (ideally quarterly figures), can serve as early warning indicators of labour market performance. Such an exercise for all

[21] For an introduction to applying such a transition matrix systematically see Schömann and Kruppe 1996.

European member states has been demonstrated, for instance, by Schömann, Kruppe and Oschmianski (1998).

3.1.3 New Jobs Dynamic

Every year, in all post-industrial societies, about 20 per cent of people will either leave a job, find a new job or do both things. Such flows into and out of work are the means by which labour markets adjust to shocks or structural change. Recently, several European countries - for instance Great Britain, the Netherlands and Denmark - have experienced continuous positive balances in those movements, in other words, spectacular employment growth within a few years. One of the questions that is far from resolved, however, concerns the quality of the new jobs being generated, including their origin (in what size of firm and in which occupations are the jobs being created, and with what employment status?). From the perspective of transitional labour markets, it is important to know whether an economy is delivering a majority of high-status, highly-paid jobs or whether recovery merely means that new labour market entrants are condemned to accepting low-status, low-paid jobs. If the latter is the case, can these low-status jobs ever provide a way into relatively highly-paid, secure jobs? In other words, is low-status entry only a transition to a sustainable and upwardly directed career or the beginning of a precarious working life?

To answer such questions it is not enough to compare aggregate stocks of employment between two points in time. Individuals should be followed, ideally quarterly, during their transitions and for a longer period of time (one to three years at least). As a result, labour market mobility patterns would emerge for certain categories of workers and age groups which could also serve as benchmarks if they were available for several countries. An excellent start in this direction was recently presented by Paul Gregg and Jonathan Wadsworth (1997). Of their many interesting findings, mention will be made of just two remarkable points about the British 'jobs machine'. Firstly, job-to-job moves on balance lead to a shift from part-time to full-time status and into clerical jobs which in later years will be the major stepping-stone for promotion to managerial positions for those who stay with the same employer. Secondly, the balance between job exit and entry suggests that job loss results in a substantial deterioration of job quality. The vacancies filled by those out of work are of extremely low quality, low paid and have a low survival rate.

3.2 Benchmarking Labour Market Policies: Possibilities and Limits

Though both employment systems analysis and the transitional labour markets approach offer analytical as well as normative devices that can be used in the development of benchmarking, the principal problems and limitations of the benchmarking method in the employment field have not been discussed yet. First of all, the issue of comparability and transferability has to be mentioned. Differences in national problem structures and institutional regimes exclude the simple imitation of best practices. This is true even for results ('product') benchmarking in the private sector. There are often no universally accepted and precise standards for LMP programmes and services (e.g. job counselling), and there is an enormous variety of programmes across the Union. As a result, cross-

national benchmarking of labour market policy will, unsurprisingly, be very difficult. If 'product' quality is to be assessed, the issue of comparability is even more critical.[22] Thus when it comes to reaching practical conclusions, the development of adequate performance indicators is as important as the inclusion of institutional differences. A wide range of evaluation techniques may be a suitable strategy for obtaining robust and valid benchmarks.

Second, data availability is a serious problem. Monitoring is still far from perfect in EU member states, because important information, for instance on long-term job prospects and future income streams after programme participation (cf. Auer and Kruppe 1996), is either not collected or is difficult to assess. However, even if the practical and technical difficulties were to be solved, there could still be severe problems with acceptance. It is far from easy to establish a consensus among member states on employment policy strategies, although recent progress has to be acknowledged (see part 4 for more details) .

A third concern is the reaction among policy-makers. It is unlikely that (cross-national) benchmarking will *quickly* induce or enhance action (programmes), reforms and competition. This is because government policy formation, unlike that of private firms, is not guided primarily by benchmarking procedures. Another reason is the limited degree of control governments can exert over economic policy. Government strategies may fail, despite the adoption of certain benchmarks, if non-governmental actors behave in a counteractive or uncooperative manner. However, to undertake a benchmarking exercise without completing the final learning stage and implementing appropriate measures would clearly constitute a failure. Benchmarking will be of very limited use and make no impact if it degenerates into a formalized process of data collection and analysis. It follows, therefore, that benchmarking should be complemented and supported by performance management systems that can provide continuous monitoring and periodic checks on results.[23] Again, it must be conceded that the recent decisions of the Luxembourg Employment Summit were a remarkable step in this direction (see chapter 4).

Despite these limitations, the prospects for benchmarking are still promising. Firstly, benchmarking could promote 'organizational learning' in governments and their agencies. Secondly, it may also in the long term encourage the diffusion of good or best practice in specific fields throughout EU member states. Thirdly, benchmarking can also contribute to the improvement of monitoring systems across the Union. However, it

---

[22] It should be noted that this is in fact an old problem in comparative politics and policy analysis. However, since benchmarking is supposed to produce concrete measures and policy improvements, these problems of comparability deserve special mention.

[23] It should be mentioned in this context that it is open to debate whether management by results or objectives should be conceived as forms of benchmarking. In spite of similarities and overlaps, we would argue that performance management (MbR/MbO) is not necessarily organized as an internal (within units of the same organization) or external benchmarking process. Hence we would prefer to distinguish benchmarking from performance management and to regards these processes as complementary rather than identical.

should be emphasized that these promises can be met only if the pitfalls and implementation errors characteristic of benchmarking (see section 1.2.3) are avoided. First and foremost, benchmarking needs a long-term perspective and commitment at all levels of the organization in question. Moreover, as already argued above, a bias towards action and experiment will be essential for success, together with the best conceivable diagnostic 'toolkit' for gathering information and monitoring success.

All types of benchmarking are applicable in labour market policy, but their usefulness and ease of implementation vary. Several options have to be considered. *Internal benchmarking* (here understood as: benchmarking within one country) will be the easiest to implement, since comparable units can be identified or operationalized without too many problems (e.g. implementation agencies, employment offices, but also regional units)[24]. *Results benchmarking* can be applied to one specific labour market programme or a specific type of instrument, as well as to programmes of the same type (e.g. wage-subsidies). *Process benchmarking* (functional benchmarking) may be more difficult. For instance, the transformation of processes into measurable performance indicators (and/or a valid process statistic) will often require thorough time studies at the work place. Even this will not suffice for the benchmarking, for instance, of case management processes initiated for the very-hard-to-place. Such processes are often based on a holistic approach that combines a set of social, psychological and skill formation measures whose quality, effectiveness and efficiency have to be assessed in their totality, not in isolation (see Schmid, Krömmelbein and Gaßmann 1994). As a rough rule of thumb, it can be said that the less standardizable a process is, the more difficult it is to measure and assess (benchmark) by use of 'hard' (statistical) data (distortion problems). These difficulties notwithstanding, process benchmarking in labour market policy can rely methodologically to a certain extent on survey research and interviews therefore being principally open to quantitative analysis.

External benchmarking (here: in an international comparative perspective) will be dealt with in greater detail in the subsequent sections. We will confine ourselves here, therefore, to a brief observation on (external) '*strategic' benchmarking*. A comprehensive undertaking of this kind would include an analysis of the long-term goals of employment and labour market policies and the programmes required to achieve those goals. It therefore comes up against all the methodological problems associated with complex cross-national comparisons, and would ideally encompass polity, politics and policies. In other words, strategic benchmarking would need a lot more research input than other approaches. Consequently, decisions on strategic benchmarking will be considerably influenced by issues of time and money. Nevertheless, strategic benchmarking could be confined to specific fields (e.g. long-term unemployment), which would probably be a more viable and frequently used option.

To sum up briefly, it can be concluded that benchmarking is *in principle* applicable to a wide range of labour market issues provided that at least four decisive conditions are

---

[24] However, control of context and intervening variables (e.g. strong supply or demand variations that may significantly influence the placement performance of employment offices) will be crucial to the validity and reliability of this kind of internal benchmarking.

met: (1) a thorough understanding of the benchmarking approach, 2) a wide ranging database and multi-method approaches, 3) adequate resources (time, money, personnel), 4) commitment of key actors involved to (organizational) learning including an 'action bias'. In other words, the requirements for the proper implementation of benchmarking are undoubtedly very high. Thus the feasibility of benchmarking in such a complex field as employment and labour market policies should also be assessed in terms of opportunity costs, i.e. the costs and benefits of alternative approaches. Moreover, it seems reasonable to expect that benchmarking will be more feasible in regions within countries rather than across countries and within the traditional (private sector) domain of benchmarking, i.e. organizational analysis. In other words, organizational entities such as public or private placement services are likely to be more easily benchmarked than policies on long-term unemployment or equal opportunities, whose success depends in part on other policies and institutions.

3.3 Benchmarking Labour Market Policy : Examples

Experience with labour market policy benchmarking is still fairly limited. However, some examples can be cited. For this exercise, we distinguish between academic benchmarking studies, on the one hand, and benchmarking as a policy device, on the other.

Research studies

From an academic perspective, Naschold and Arnkil (1997) conducted a „benchmarking" exercise in which the performance of a single local employment office in six countries (Sweden, Norway, Denmark, Finland, Great Britain, New Zealand) was scrutinized and compared with reference to „best practice criteria" formulated from a new public management perspective. Although this study provides some useful (qualitative) evaluation criteria for employment offices, it does not discuss in detail the still controversial issues of the 'right' placement strategies and the possible shortcomings of a strong market orientation (e.g. creaming effects). Thus the benchmarks developed remain, at least in part, ambiguous and even inadequate as far as labour market performance criteria are concerned.[25]

Mention should also be made of two Dutch benchmarking studies.[26] One of them addresses Dutch competitiveness, which is defined "as the ability of the private sector and government to generate prosperity and jobs in a highly competitive environment" (Dutch Ministry of Economic Affairs 1995: 3). The Netherlands is benchmarked against Belgium, Denmark, Germany, Japan and the United States in the areas of infrastructure, monetary and fiscal stability and education, as well as capital, product and labour markets. Simple ordinal scales (high, average, low) and/or statistical comparisons and cross-national ratings are used to illustrate core performance dimensions. The labour

---

[25] For a more comprehensive framework of labour market performance criteria see Schmid 1996a. See also the documentation of the debate on placement services and the long-term unemployed in Schmid et al. 1997.

[26] These studies fall between our two categories of research and policy device; we present them here under 'research benchmarking' as they are not directly related to policy 'action'.

market performance criteria are labour productivity, days lost through strikes, wage costs, labour market dynamism, wage dispersion, long-term unemployment and utilization of labour potential. It is obvious that this study addresses labour market performance rather than policies. The approach pursued is very 'pragmatic' insofar as neither the pros and cons of different PIs nor crucial institutional differences are discussed in any detail. In consequence, few if any inferences about policy can be drawn from this study. However, more exhaustive follow-up studies have been announced (ibd.: 21).

The second study assesses the impact of Dutch welfare state policies on competitiveness relative to the same countries used in the first study, plus Sweden and the UK (Dutch Ministry of Social Affairs and Employment 1996). It includes income distribution, health care, occupational health and safety, social security as well as the labour market and industrial relations. Methodologically, it resembles the first study in its use of ordinal- and ratio-scaled ranking schemes and summaries. However, the institutional analysis digs deeper and the compilation of performance indicators is better reflected.

On the whole, these two studies highlight the growing importance of international comparisons and benchmarking in times of higher competitive pressures and globalization. This challenge should be met with theory-driven research design and thorough operationalization.

Benchmarking as an explicit policy tool of labour market policy

In Europe, the United Kingdom makes use of several forms of benchmarking in labour market policy. In the area of quality standards, for instance, the British employment service has been awarded the '*Investors in People*' quality certification, which recognises that the organization has met certain educational and further training standards. This program may serve as an example of quality competition between different public sector organizations. Another example of benchmarking is the *Jobseekers's Charter*, the employment service's implementation of the *Citizen's Charter* initiative (Doern 1993; Pollitt 1994). The *Jobseeker's Charter* sets out what kind of services jobseekers can expect from the ES.[27]

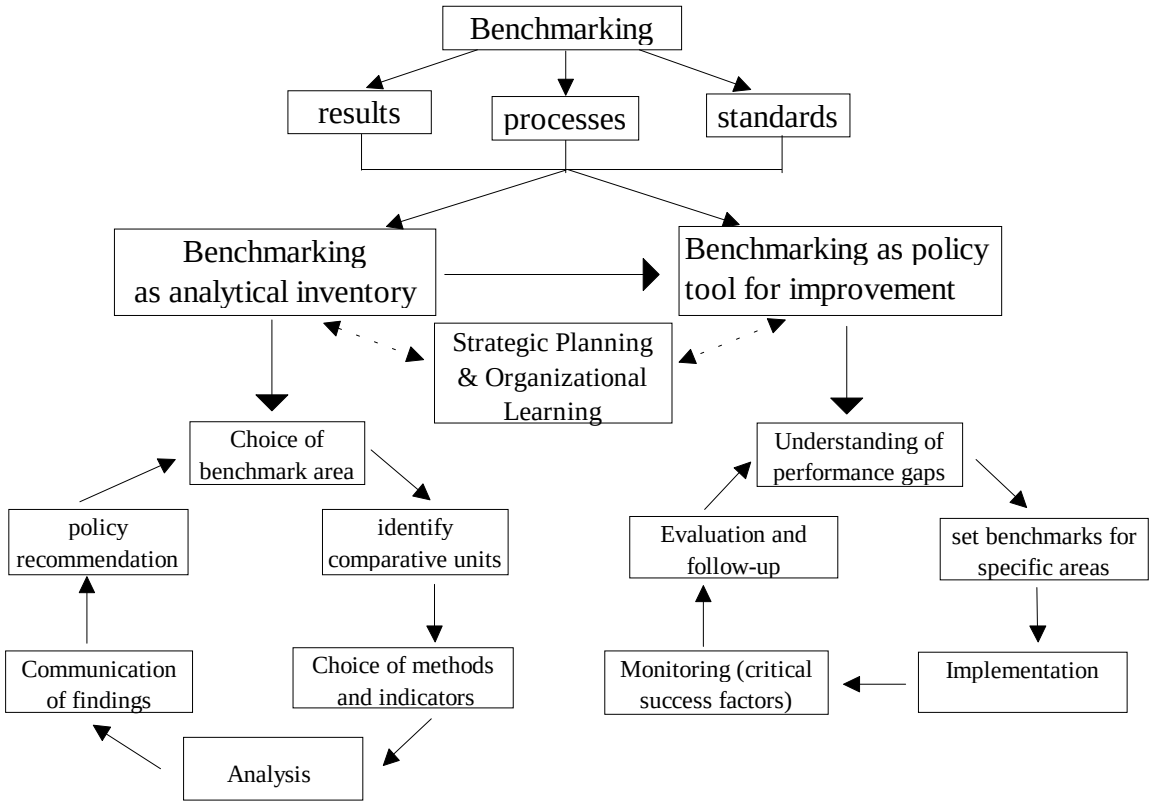
3.4 Interim Summary: Benchmarking in Labour Market Policy

As already noted "benchmarking" in labour market policy includes pure evaluative studies on the one hand and use as a policy instrument on the other. From a theoretical point of view, benchmarking analysis should always be related to and serve as a basis for policy action designed to bring about positive changes, as emphasized in section 1.2.3. In other words, analysis, learning and action are highly interdependent in the benchmarking concept. In our opinion, this insight from the private-sector benchmarking literature

[27] Examples are the guarantee of a personal interview at the first claim, the laying down of maximum waiting times for service or a response to a telephone call and a guarantee that job-seekers will be sent the right money, on time.

holds true for public-sector benchmarking as well. In this strict sense, genuine research studies should not be described as benchmarking exercises. However, Figure 5 is an attempt to summarize both the theoretical objectives and the empirical findings of benchmarking applied to the public sector in general and labour market policy in particular. From a more pragmatic point of view, the useful definition of benchmarking proposed by the Commission should also be kept in mind: "In the field of employment, benchmarking is a formalized process by which employment performances of different countries are compared with each other in order to highlight the best performing ones, to set global targets for progress in the employment situation over the medium/long-term and to identify which policies have been most effective in raising the level of performance" (Joint Employment Report 1997). In the next part of our paper, we look at benchmarking in four concrete fields of the European Employment Strategy.

Figure 5: Benchmarking in Public and Labour Market Policy



4. Benchmarking the European Employment Strategy: Some Theoretical and Practical Considerations

The Luxembourg Employment Summit (November 1997) and the new Treaty of Amsterdam (though not yet ratified) represent substantial steps towards effective coordination of employment policies at the European level. Though it can be objected that, given the scale of unemployment in Europe, much more needs to be done, it must also be recognized that the new employment title of the Amsterdam Treaty translates the programmatic statements of the Essen Strategy into legal obligations. This has been underlined by the Luxembourg decisions and the 19 employment policy guidelines. The higher profile of employment and social policy issues at European level is reflected particularly in the three key decisions taken at the Employment Summit that require member states:

- to commit themselves to a national surveillance procedure, to translate the annually endorsed employment policy guidelines into National Action Plans (NAPs) and to report on the implementation of NAPs;

- to create a performance management system (MBO) by defining quantified targets for some of the guidelines (currently for three, namely the guarantee that any young person or adult will be offered a place on an employability scheme before they have been unemployed for six or twelve months, the commitment to increase the number of active measures for the unemployed by at least 20% in order to promote activation, which should be gradually reaching the average level of the three most successful member states)

- to set up a convergence process by fixing objectives for the employment level in Europe.

Since the implementation of these decisions requires substantive analytical work, the European Union has initiated a process of benchmarking of employment and labour market policies. There is now a good chance that the dual function of benchmarking as an analytical inventory and a policy tool for (continuous) improvement will be implemented 'in real-life' across the Union. The benchmarking of labour market policy is not just fashionable, it is also extremely relevant to both academics and policy-makers in Europe. This is confirmed by the fact that DG V of the Commission had started to work on the benchmarking of labour market policies before the Amsterdam and Luxembourg decisions were taken, as is documented in the Joint Employment Report 1997 (Commission 1997a) and in three analytical papers on the benchmarking of long-term unemployment, youth integration and the gender dimension of equal opportunities policies (Commission 1997b, c, d). In addition, the RESEARCH Network was instructed, within the framework of the European Employment Observatory, to prepare a report on benchmarking labour market performance and policies in 1997. The following observations on the three policy areas of long-term unemployment, youth integration and equal opportunities are complementary to the aforementioned work.

The section is organized as follows. In accordance with the target-oriented evaluation approach (Schmid, O'Reilly and Schömann 1996; 1996a), theoretical issues relevant to each policy objective will be briefly considered by way of introduction to a discussion of measurement problems and the development of pragmatic performance indicators for benchmarking. Youth unemployment is dealt with in somewhat greater detail in order to illustrate the pitfalls and difficulties of benchmarking. Our main focus will be on measurement issues.

4.1 Long-term Unemployment

4.1.1 Theoretical Issues

The fight against long-term unemployment is one of the EU's priority employment policy objectives (EU 1995), as befits the seriousness of the problem (EU 1996a: 95-97). The causes of long-term unemployment (LTU) are still a matter of controversy. At one end of the range of views, labour market institutions (e.g. long-term and generous unemployment benefits) are seen as the scapegoat, while at the other extreme LTU is regarded primarily as a selection problem (creaming of the fittest, exclusion of those with cumulative risks) in times of demand-deficient mass unemployment (see Heise 1995). Nevertheless, it must be noted that LTU in the EU affects not only the hard-to-place, but increasingly also people with skills and formal qualifications. Moreover, the problem structure varies across the Union: in some countries, long-term unemployment is mainly a problem for the young, in others for older workers. However, any comprehensive attempt to benchmark long-term unemployment would first have to deal with the possible multiplicity of causes before good or best-practice remedies could be benchmarked. In this sense, a comprehensive institutional analysis would have to cover the following labour market institutions:[28]

- the wage formation system
- the social security (especially unemployment insurance) system
- the tax system
- the educational system
- (targeted) active labour market policies, especially training measures and comprehensive schemes.

Such an analysis would also have to address the role of general employment policy, above all fiscal and monetary policy. Examination of the relevant institutions and policies should always be the starting-point of any benchmarking activity. Secondly, the policy options for lowering long-term unemployment have to be considered. Here, the main distinction is between demand and supply-oriented measures. According to the OECD (1992: 25), measures can be divided into three groups:

---

[28] The debate on LTU is more or less driven by neoclassical assumptions focussing on the relations between minimum wages, and/or social security levels as an equivalent, wage differentiation and (the duration of) unemployment. For empirical studies with such frameworks see Bellmann 1996; Rolle and van Suntum 1997. For different theoretical and empirical findings see, among others, Schmid and Reissert 1996. A recent discussion of the so-called poverty trap can be found in OECD 1996a.

- those intended to encourage, improve and assist in job search;
- those intended to reactivate and improve the skills of the long-term unemployed;
- those intended to improve options for the long-term unemployed by reducing their costs to employers, e.g. wage cost subsidies, encouraging self-employment, paying individuals subsidies to encourage them to accept jobs they might not otherwise consider, and direct job creation schemes.

A fourth option is also mentioned, namely the 'focused approach', i.e. measures targeted on specific groups within the long-term unemployed population. From an international comparative perspective, of course, these principal options cover an enormous range of measures, to say nothing of EU activities within the ESF.[29] The question of how benchmarking is to deal with this complexity is an interesting though as yet unresolved issue. We now turn to measurement issues.

4.1.2 Measurement Problems

The central policy goal is to minimize LTU, ideally to eliminate it altogether. This will of course be a medium or long-term goal for most countries under the present conditions. However, any policy strategy that seeks to lower LTU significantly over the medium term will have to address measurement issues right at the beginning in order that adequate monitoring and evaluation systems can be put in place.

Firstly, the usual stock data compiled for administrative records are prone to underestimate the real scale of long-term unemployment. One important reason for this lies in the fact that in (alternative) survey research respondents report themselves as long-term unemployed (for 2, 3 or more years) even though they have briefly interrupted their unemployment 'career' on one or more occasions. This kind of 'punctuated long-term unemployment'[30] is not captured by official statistics, which count the persons concerned as new entrants into unemployment. In addition, account has to be taken of discouraged workers, who do not usually appear in official statistics. Although (matched) labour force survey data may have deficiencies, such as sample attrition, inconsistent answering or recall bias (see OECD 1995a: 26), this kind of data is in theory preferable to administrative records. Moreover, the incidence of long-term unemployment (LTU as percentage of total unemployment) - as it is annually presented in the OECD Employment Outlook - is calculated on the basis of the European Labour Force Survey for most member states and can thus be considered as the best possible measure for EU comparisons.

Another problem concerns the usual way of dividing the unemployed population into the short-term (< 1 year) and long-term (> 1 year) unemployed (cf. Karr 1997). This method counts the 'indefinite' duration, since the actual period of unemployment for any one individual is not yet known. This means, on the one hand, that the actual volume of unemployment cannot be precisely calculated and that, on the other, it is impossible to

---

[29] For a brief overview across the Union (EU-12) see Gaß, Krömmelbein and Schmid (1995). For LMPs targeted on the hard-to-place see Erhel et al. (1996).

[30] For a thorough German study on this subject see Büchel 1992.

identify all those who subsequently enter long-term unemployment (after days, weeks or months). In other words, recording the actual shares of short-term and long-term unemployment is useful only if it is done after the end of all unemployment spells. The analysis of outflow cohorts which comprise the completed duration for any single person is an equivalent for this and can be used to calculate the exact volume of LTU (number of unemployed > 1 year x duration) + (number of unemployed < 1 year x duration = total volume in days). Calculated in this way, the LTU share also provides information on the stock of those already classed as long-term unemployed at the reference date and on those joining that category on that date. Another implication of the standard (German) counting system is that the volume of unemployment is reduced for a year (364 days) as the duration is only counted from day 365. This is the decisive factor in the underestimation of the volume of long-term unemployment in the official statistics. These aspects should be borne in mind when analysing long-term unemployment.

However, decomposition of the (long-term) unemployment volume into flows and duration should be a principal standard in order to detect, for instance, 'duration distortions' in total unemployment rates. Moreover, although the relationship between the total unemployment rate and the incidence of long-term unemployment or the LTU rate is not very strong across countries (notable exceptions notwithstanding), this gives us valuable information on national patterns (see OECD 1993a: 86f; 1995a: 20-25). Lastly, in order to relate LTU to labour market dynamics, it would be desirable to estimate hazard or transition rates from short-term to long-term unemployment.

4.1.3 Performance Indicators

Benchmarking should aim to use performance indicators that are both concise and relevant. Following on from the previous considerations, desirable performance indicators of long-term unemployment would be:

- total volume of unemployment, expressed (in days or weeks) as the product of completed durations and the number of unemployed ;
- distribution of this volume between long-term (> 1year) and short-term unemployment (< 1 year);
- hazard or transition rates (probability measures) from short-term (<6 or 12months) to long-term unemployment (>6 or 12months)[31]

Though not a performance indicator in the strict sense, it would also be desirable for the conducting of detailed and in-depth studies to gather the following data (from which benchmarks might be derived for specific issues):

- decomposition of the long-term unemployed by completed durations (at the reference day of a survey) and group characteristics such as age, sex, education, occupation, specific placement impediments.

---

[31] This will be an important indicator for the performance management of the quantified targets on employability.

However desirable such information and indicators might be, it is impossible to ignore the realities of data availability and the compromises that inevitably have to be made. The following indicators of long-term unemployment are both indispensable and available:

- incidence of long-term unemployment (as share of total unemployment), as reflected in labour force survey data or a combination of these and administrative data;
- long-term unemployment rate (as percentage of the labour force);
- the rate of change (as percentage of previous period) in these two indicators.

These indicators mainly address the 'diagnostic' side of the coin. From an evaluation perspective, the ALMP effectiveness of measures targeted on the long-term unemployed also needs to be addressed of course in order to attain benchmarks for policy programmes:

- follow-up employment rates after programme participation (placement immediately after programme end; 1-year, 5-year follow up);
- income levels of programme participants after re-employment.

Again, these are undoubtedly desirable indicators that are not available for all (targeted) measures. However, it seems possible to construct the following indicators for all EU member states in order to get at least a proxy for the relevance of policies to the fight against LTU:

- percentage share of (previously) long-term unemployed people in ALMP-programmes;
- placement or inflow rates of (previously) long-term unemployed people into ALMP-measures and/or regular employment.

4.2 Youth Employment and Unemployment

4.2.1 Theoretical Issues

There are two main approaches to tackling youth unemployment in the European Union (EU 1995a: 114):

- encouraging young people to stay in education and initial training longer in order to minimise the number of people entering the labour market without adequate basic qualifications;
- facilitating the transition from school to work by developing closer links between the two and by improving access to training for the most vulnerable.

The policy objectives related to this field are complex. Two priorities are *sustainable labour market integration*[32] and the fostering of *equal opportunities*. Of growing

[32] In this context, this means that educational and labour market policies should be future-oriented, i.e. aquired skills and competences should help workers to find employment after the spells of unemployment they will inevitably experience.

importance are the goals of *employability*[33] and the *prevention of mismatch* (see Ryan and Büchtemann 1996, for a thorough discussion). The array of specific policy options mean that priorities (goal hierarchy) have to be established before any benchmarking can begin.

The youth unemployment problem is a matter of labour market as well as of education policy. Performance will presumably be improved if these policy areas are closely coordinated, in terms of both objectives and measures. Labour market policy will not be able to compensate fully for structural deficiencies in education systems. Strategic benchmarking, in particular, would have to address these issues. In sum, any attempt at benchmarking in the sphere of youth employment and unemployment has to consider at least three interrelated systems:

- *the primary and secondary education system*, which acts as an initial filter, structuring career options and employment prospects;
- *the vocational training system*, which acts as a 'second filter' for most school leavers and determines the quality of skills and competences and their adaptability to changing labour market and workplace requirements;
- *the labour market policy system* with active measures targeted on young people, particularly on those with cumulated risks of social exclusion; youth measures may also provide 'second chances' for young people for who have failed to enter the labour market through the 'two-filter system'.

Vocational education and training is a subject that exemplifies the risk of false conclusions being drawn if too much faith is placed in statistical data. In statistical terms, Germany is undoubtedly the best performer in youth unemployment. Nevertheless, serious problems and specific institutional requirements related to the 'dual system'[34] should warn us against excessive enthusiasm for this specific training approach . The starting-point for benchmarking vocational training must be the recognition that the link between school and firm-based training is configured in at least three different ways in OECD countries (cf. OECD/Céreq 1994: 42-49):

- in the German-speaking countries, the dual system is prevalent;
- in some countries, school-based training predominates but considerable efforts have been made to develop various forms of alternating firm and school-based training ;
- in other countries, the creation of a vocational training system and alternating training between school and firm within such a system is currently being debated (see Figure 6).

---

[33] Due to the fact that there is now a growing number of young people aged 20-25 in the EU who are neither employed nor in any kind of education or training (EU 1995a).

[34] To name just a few: the lag in the adaptation of occupational profiles to future needs, uneven distribution of employers providing apprenticeships (free-rider problem), political disputes on the coordination of school curricula with workplace requirements and a growing mismatch between supply and demand of apprenticeships, especially in rural East-German regions (see, among others, Franz and Soskice 1994; Wagner 1998)

Figure 6: Vocational Training and Education Systems in Europe

Dual system approach[35]	School-based and alternating approach	Relative Openness: Institutional Choice
Austria Germany Switzerland (non-EU member) -------- Portugal ←Denmark→	Belgium Finland France Greece Italy Netherlands Spain Sweden	Ireland United Kingdom

An obvious option would be to start by benchmarking countries that have the same approach. Subsequently, countries with different institutional frameworks could be compared in order to identify typical policy patterns. For instance, job guarantee programmes for unemployed young people can be found only in countries with an alternating training system (Belgium, Finland, Ireland, Netherlands, Sweden and Denmark). Thirdly, institutional frameworks could be matched and compared with a requisite variety of instruments in order to identify functional equivalents. This would be an ideal benchmarking process that would avoid misleading conclusions.

4.2.2 Measurement Problems

Which indicators are essential in this area? For the purposes of benchmarking, the first thing to be decided is whether to consider employment and unemployment levels for the 15-25 age group as a whole or whether to divide the group into two sub-groups, 15-20 and 21-24. As there are no serious data restrictions, the general preference will be the second option since, these two groups represent different stages of labour market transition. Secondly, a denominator has to be chosen that reveals significant differences in performance. For instance, if the unemployment rate is calculated as a percentage of the labour force (in the relevant age group), Spain, Italy, France, Belgium, Ireland and Greece emerge as particularly poor performers. But if youth unemployment is measured as a proportion of the age group as a whole, the results are different: in the teenage group (15-20 years) France and Belgium are second-best performers behind Germany, while the UK has the most severe unemployment problem after Spain.[36] The distortion effects in this age group reflect the higher share of those who are still in initial education

---

[35] Though Portugal is obviously not a German-speaking country, it does have a dual apprenticeship system that is strongly institutionalized through a tri-partite arrangement between the social partners, who play an important role in regulating the system. Denmark is a special (intermediate) case, since it has a long tradition of alternating school and on-the-job training like Germany, but practical training is often provided in school workshops due to a lack of company apprenticeships.

[36] For those aged 20-24, however, the results of countries' performance do not change structurally (though the quotas are, of course, lower) when using the population as denominator (EU 1995a), since the educational system does not interfere so much as in the case of the teenage group.

and/or economically inactive (cf. EU 1995a: 66-71). In the case of youth unemployment, therefore, the population seems to be a better denominator than the labour force.

For a positive benchmark of youth integration, it would be desirable to sum the number in the regular labour market and those in any kind of education or training. By relating this sum to the age group as a whole, we would get a fairly good indicator of the share of young people who are economically active. However, since the labour market situations of young people will change quite rapidly in a number of countries, the intervals between sample surveys should be short in order to increase the validity of the proposed numerator, which has to be averaged. Unfortunately, at present, this indicator cannot be calculated for some member states.

But even if we had such an activity indicator for young people, it would tell us nothing about the quality of their jobs, education or training measures. For assessing the quality of young persons' labour market integration or status, educational indicators will still be the best proxy measure. In this regard, the OECD suggests a couple of useful indicators (OECD 1995b: 214):

– share of young people successfully completing a full course of secondary education (private or public) at normal age by type of educational programme.

In the OECD statistics, this indicator relates to general upper seccondary education, as well as to vocational training and apprenticeship. It is also useful to relate unemployment to education (cf. ibid.: 227-31):

– unemployment rates by level of educational attainment (here to be adapted to ages 15-19, 20-24, both sexes);
– additional option: the range of unemployment rates by level of educational attainment (to be adapted to ages 15-19, 20-24, both sexes).

These indicators encompass primary and lower secondary education, upper secondary education, non-university and university education. The latter indicator is illustrative for international comparisons. Finally, there is also an indicator measuring the labour force status of those leaving the education system (ibid.: 248f):

– unemployment rates after leaving education for leavers from different school levels and unemployment rate in the total labour force (in corresponding age group), to be measured e.g. after 1 year and 5 years.

These OECD suggestions are promising indicators for benchmarking activities in general, not just those relating to young people.

4.2.3 Performance Indicators

In sum, we suggest the following set of indicators:

– Unemployment rates by level of educational attainment (to be adapted to ages 15-19, 20-24, both sexes); same for employment rate (and by occupation, together with sectoral distribution);

– the rate of change (as percent of previous period) of these indicators;

– share of young people successfully completing a full course of secondary education (private or public) at normal age by type of educational programme. We also propose the following as an additional indicator that is desirable though not available:

– share of the 15-24 age group in employment or in education/training (comprehensive positive integration indicator).

Moreover, as in the case of long-term unemployment (see above), a basic indicator set should also include:

– hazard or transition rates (probability measures) from short-term (<6 or 12 months) to long-term unemployment (>6 or 12 months);
– follow-up employment rates after participation in active measures targeted on young people (placement immediately after programme end; 1 year, 5 year follow up),
– income levels of programme participants after entering employment.

4.3 Equal Opportunities[37]

4.3.1 Theoretical Issues

There is a wide range of EU activities in this field that are now summarized for the first time in the 1996 annual report of '*Equal Opportunities for Women and Men in the EU*' (EU 1997).[38] It is again beyond the scope of this paper to discuss all these issues from a benchmarking perspective. Rather, we concentrate once more on measurement problems. Firstly, however, it should be remembered that comprehensive benchmarking of this subject would have to take into account quite a number of policy fields. These would include the following at least:[39]

– wage policies,

– working time policies,

– labour market policies targeted at women,

– parental leave policies,

– childcare policies,

– tax and social security policies.

The linkages between these policies are of crucial importance in any attempt to reduce gender inequalities in labour markets. Quick-fix solutions will not, therefore, be possible.

---

[37] This discussion of the benchmarking of equal opportunities focuses solely on the gender dimension of labour markets and excludes subjects such as, for example, the labour market situation of disabled persons or ethnic minorities.

[38] It is intended to publish such a report every year from now on, in order to strengthen the monitoring of policy progress. The 1996 report indicates, for instance, that an important benchmark will be to what extent EU directives have been already implemented by member states.

[39] Among the vast literature see in particular Humphries and Rubery (eds.) 1995; O'Reilly 1996; Rubery, Fagan and Maier 1996; Fagan and Rubery 1996; Beckmann (ed.) 1996.

Moreover, many aspects of the aforementioned policies still have to be adequately researched. For instance, it is only relatively recently that the question of whether part-time work is a bridge or a trap has been a subject for research. From the point of view of transitional labour market theory, it is important to ascertain whether part-time employment offers adequate labour market entry, whether it enables transitions into full-time work (and vice versa) and whether part-time employment can represent a qualitative alternative to full-time employment over the long run without individuals incurring a risk of downward social and labour market mobility or exclusion (O'Reilly and Bothfeld 1998). To address this question, it is necessary to study the empirically ascertainable patterns of transitions between different employment statuses. At present, however, there is little empirical data on this topic. Hence it follows that it is far too early to expect any worthwhile cross-national benchmarking in this area.

4.3.2 Measurement Problems

The prevention or mitigation of wage discrimination, labour market segmentation and occupational segregation are the core objectives of equal opportunities policies. We will confine ourselves here to a few brief comments on the measurement of segregation and the gender-pay gap.

A traditional way of measuring gender segregation is to use summary index measures (hereafter, we draw largely on Rubery, Fagan and Maier 1996). Disaggregation by (standardized) occupations is the best method of capturing the characteristics of segregation. However, three problems with index measures have to be considered. The first is how to construct the index in order to measure changes in segregation when there are simultaneous changes in the employment structure and women's share in the labour force. Secondly, the choice of occupational classification scheme is also a crucial issue. Most national classifications do not fulfil the requirements of comparability, and there are good reasons for doubting the validity and reliability of international harmonized classifications (cf. also Hakim 1996: 70). The third problem is that index measures tend to hide divergent trends, since they do not indicate those labour market segments in which segregation is increasing or those in which it is decreasing (simultaneously). However, if such indices are to be used, then the index of segregation[40] seems to be a favourable measure as it is sensitive neither to the female share in the labour force nor to changes in the size of sectors in which women are over- or underrepresented. In contrast, the dissimilarity index (also known as the difference index), the most widely used measure of occupational segregation, is sensitive to these two dimensions because of the way in which it is constructed.[41]

---

[40] The index of segregation measures the relative share of the female labour force in 'male' and 'female' jobs; a cut-off between these gender categories is fixed where the number of women and men employed in 'female' jobs equals the number in the total female labour force.

[41] The dissimilarity index expresses half the sum of the absolute differences between the male and female coefficients of representation in each occupation; the dividing line between 'female' and 'male' occupations is based on the over or underrepresentation of women in relation to the total share of women in the labour force.

Measurement problems also affect attempts to measure the gender pay gap as well. These problems have their origin, on the one hand, in the availability[42] and validity of wage data and, on the other, in the methodological difficulties of measuring the impact of legislation on equal pay and equal opportunities.

4.3.3 Performance Indicators

With these limitations in mind, we suggest the following indicators as a basis for benchmarking:

- labour force participation rate in the core age group (25-54) as a central indicator of female labour market integration;
- decomposition of the former indicator into employment and unemployment rates;
- female full-time and part-time employment rates, as shares of the female population of working age; this gives a far more differentiated insight of female integration than the labour force participation rate;
- female wages compared to male wages (e.g. mean annual incomes; hourly wage rates in industry or services), for age groups and by educational attainment (gender pay gap indicator);
- the index of segregation and (for reasons of wide availability) the index of dissimilarity.

4.4 Employment Creation and Active Labour Market Policy

A key focus of active labour market policies (ALMP) is the (re-)integration of target groups (e.g. women, long-term unemployed etc.) into the regular labour market through job creation measures. Hence changes in output and outcome indicators related to these groups could be possible benchmarking indicators. A key question is whether ALMP leads to the net creation of jobs. In an aggregate impact analysis framework, active labour market policy is expected to affect wage negotiations (Bellmann and Jackman 1996). For instance, the aggregate wage level can be assumed to be dependent on the extent of subsidised employment (Calmfors and Skedinger 1994). In theoretical terms, the impact on employment is uncertain. On the one hand, positive effects can occur if it is assumed that LMPs (especially training measures) help to generate more competitive human capital. The resulting competitive pressure on 'insiders' (job-holders) will reduce the aggregate wage level, thereby producing a net increase in employment. On the other hand, these positive effects could be countervailed by possible negative effects such as free-riding, substitution and dead weight effects. Empirically, these effects can be illustrated with indicators of regional or sectoral labour market dynamics, e.g. changes in market processes after the implementation of a measure.[43] Another option for measuring the aggregate employment impacts of active labour market policies is to use changing labour market dynamics as an indicator. Longitudinal survey data is well suited to this purpose. For example, there is a question on occupational status at time t and time t+1 in

---

[42] EUROSTAT may soon provide more recent and comprehensive data in the new Earnings Statistics.

[43] See for instance Anxo's (1997) study on recruitment subsidies in Sweden, which reported significant dead weight effects.

the European Labour Force Survey. The answers to this question can demonstrate how employment dynamics in target groups change during or after programme implementation.

National ALMP mixes, which are an important part of employment regimes, should be explicitly taken into consideration in any analysis of job creation. As with vocational training, there are three different models in the EU:

- The main feature of the *'continental model'* is the creation of jobs in an often semi-governmental, non-market sector as a means of facilitating reintegration into regular labour markets. One problem with this strategy is that it may lead to "recurrent unemployment" rather than reintegration (Büchtemann and Brasche 1985).
- The *'Scandinavian model'* places considerable emphasis on the creation of additional permanent jobs in the public sector (Esping-Andersen 1994). The shortcomings of this approach lie, among other things, in high costs and saturation limits. Furthermore, it has a positive effect on labour force participation (particularly women) to the possible disadvantage of the unemployed.
- In the *(British) market model*, subsidized employment creation does not play a significant role in labour market policy. Moreover, since the neo-liberal regime of the Thatcher years, active and passive labour market policies have been severely cut. The main burden of job creation falls on supply-side policies such as tax incentives and reductions in employers' social security contributions.

In sum, a number of different indicators will be required for a thorough analysis of net employment effects. It is not possible directly to measure the net employment creation effect of ALMP. In consequence, proxy indicators such as transition rates and changes in stocks or durations of unemployment will have to be retained.

This leads us to the final conclusions of this chapter. The challenge of benchmarking is to find policy mixes that seek to balance the goals of overall employment creation, labour market integration of the unemployed and target groups and equal opportunities, as well as structural change, innovation and competitiveness. This task cannot be achieved by adopting a single monolithic approach: even viewed in isolation, each of the fields and policy objectives under consideration here is extremely complex, to say nothing of the interdependencies between individual policy strands. To be viable, therefore, a benchmarking strategy would have to start by detecting and explaining broad performance and ranking patterns by using a basic set of indicators. Once the foundations had been laid in this way, more specific and detailed benchmarking analyses could be carried out, with adequate theoretical frameworks being provided by the employment regimes, target-oriented policy evaluation and transitional labour markets approaches. Theory aside, the new direction taken by the European Employment Strategy clearly defines the principal sub-areas that analytical benchmarking will have to address in the next few years. The next chapter introduces a benchmarking instrument that can be used to begin the necessarily long process of devising differentiated benchmarking activities.

5. The Radar Chart Approach as a Benchmarking Tool: Concept and Application

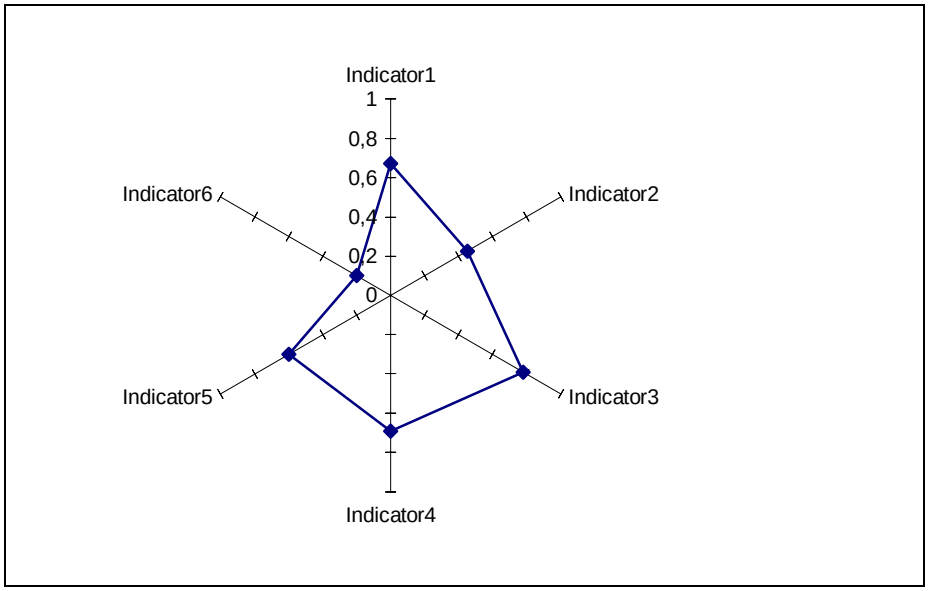
It has been argued above that a thorough benchmarking of labour market performance and policy would require a wide range of information and evaluative methodologies. Only a comprehensive approach can avoid the shortcomings of more limited analyses and ensure that the lessons for policy-making, which are the ultimate goal of benchmarking, are well and truly learnt. However, given the resource constraints in the social sciences and in politics, it is clear that such comprehensive studies will be only rarely undertaken. One way to support such ideal-type studies is to develop tools that can help to reduce the complexity without losing relevant information. It is our view that radar charts might be just such a tool.

Thus the following section is given over to an examination of this specific method, which could be an important element in the "toolkit" required for comprehensive benchmarking studies. We will seek to demonstrate that radar charts have several advantages of index building as well as a wide range of possible applications. However, it will also become clear that benchmarking cannot be based exclusively on radar charts, since they are ultimately unable to explain the reasons for performance gaps.

The principal method of using radar charts is already well established in economics and in private-sector management (see among others Albach and Moerke 1995; Bogan and English 1994; Domptin 1997). To the best of our knowledge, however, it has not yet been applied to labour-market performance and policies. At first sight, the charts in question seem to offer nothing more than four or more performance scales presented in one integrating radial chart which looks similar to a radar screen or a spider-web, hence the designation 'radar chart approach'.[44] Alternatively, we can also speak of the 'SMOP-approach', for the following reason. Connecting the performance levels attained in each dimension of the radar chart by straight lines produces an angular plane figure (see Fig. 7). The surface area of this figure can be calculated to give a (dimensionless, abstract) mathematical expression of the overall performance achieved in all measured dimensions. The abbreviation SMOP is derived from this feature, standing for 'surface measure of overall performance'. Moreover, since the SMOP represents a calculable index measure, it should be clear that the radar-chart-approach has more to offer than just smart graphical presentations of certain performance results.

[44] In fact, there is as yet no scientific definition either of the chart or of the relevant mathematical method, which might be due to the fact that many or most applications are confined to the descriptive illustration provided by the chart.

Figure 7: The Radar Chart (Surface Measure of Overall Performance, SMOP)



The section is organized as follows. Firstly, we outline the principles of the SMOP approach, sketch out some possible applications and discuss some of the related methodological issues. Secondly, a number of applications of the SMOP-approach will be demonstrated, with a brief examination, firstly, of a comparison of performance in long-term unemployment, youth integration and equal opportunities policies across selected countries and over time and, secondly, of national public employment services (PES).

5.1 The Principles of the SMOP Approach

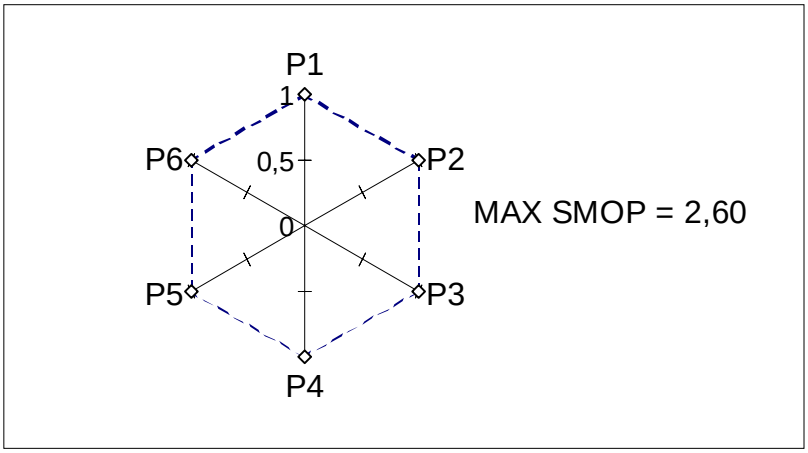
The radar chart approach has four main goals. The first of these is the visualization of interrelated performance measures through standardized scales. The second and primary goal is to produce a very effective and revealing description of selective performance dimensions in *just one* synthetic indicator. The surface of the radar chart can, therefore, be understood as an illustration of the performance of the labour market system, provided that enough policy or performance dimensions are included. Thirdly, the change in the overall performance between two points of time can be analyzed. The increase (or decrease) in the arithmetical sum of the surface indicates the improvement (or deterioration) in total performance independently of countervailing effects (improvement here, deterioration there) that might possibly have taken place. Fourthly, the shape of the radar chart as well as the overall performance measure can be used for comparisons of private firms, public agencies or countries.

In our examples and applications, the scales of each of the axes included in the polygon are standardized and have the maximum value of 1 (or 100%)[45] which

[45] This is not necessarily the case, and different scales within one chart are possible (see e.g. Domptin 1997). In this case, however, the chart loses much of its illustrative attractiveness as the different sub-units of analysis can no longer be compared directly.

corresponds to the theoretical maximum performance of the measured dimension. In other words, if all axes had a value of 1, the theoretical benchmark (best performance) would have been achieved and the surface would attain its maximum area (see Figure 8).

Figure 8: Radar Chart with Theoretical Maximum Performance



The following formula is used to calculate the surface measure (SMOP):
SMOP = ((P1*P2)+(P2*P3)+(P3*P4)+(P4*P5)+(P5*P6)+(P6*P1)) * sin 60°/2, or generally:
SMOP = ((P1*P2)+(P2*P3)+(P3*P4)+(P4*P5)+(P5*P6)+....+(P*n*P1)) * sin (360/*n*)/2.

This formula makes it clear that SMOP is an additive index measure. As such, it is beset by the methodological problems typical of all index measures. These are (Scheuch and Schimmelpfennig 1974):

- the theoretical deduction of indicators from theoretical terms ('rules of correspondences');
- the optimal selection or choice of indicators;
- the combination and weighting of indicators;
- the validation of indicators/indices.

In addition, the SMOP approach shares the specific assumptions of *all* additive indices, which have long been subject to criticism. In our case, there is no problem with respect to interval-scaled data, which is one of the necessary conditions for constructing additive indices. However, a second assumption of additive indices also applies to the SMOP, and therefore requires further attention. There is an implicit assumption that low values in one dimension can compensate for high values in another. This is, principally, an empirical hypothesis which cannot be verified solely by the a priori assumptions built into the model. However, as will soon become clear, we are using the SMOP as an index of observable performance outputs[46] which can justifiably be deemed to be of equal importance. Our SMOP is without doubt a theoretical construct rooted in a specific scientific and normative point of view. It does not reflect, for instance, the policy preferences of any of the governments of those countries we have addressed. We do not

[46] And not, for instance, for measuring social attitudes or so for which the aforementioned objection is far more important.

consider this to be a disadvantage, since we explicitly state what we are measuring. However, we do admit that some of the problems relating to the actual interactions and interdependencies between the indicators used are still unsolved. This is due to problems of data availability which forced us to use 'second-best' indicators. Thus some dimensions remain underrepresented, and parts of the indicators still overlap. Moreover, the issue of adequate weighting remains controversial. This will be discussed in more detail below. On the other hand, as has already been mentioned, the radar chart is well suited to indicating trade-offs between policy goals. The assumption of compensating effects is hence an explicit feature of its construction and therefore justified. Although the radar chart approach or SMOP is currently at a very early stage of development , we believe that the unresolved problems can be overcome. The discussion that follows should be regarded as the preliminary outline of a prototype designed to establish radar charts as an instrument in the benchmarking of labour market performance and policy

5.2 Measuring Labour Market Performance and Policies: Prototypical Applications of the Radar Chart Approach

In this section, we will demonstrate how the SMOP approach can be used in the benchmarking of long-term unemployment, youth unemployment and the gender dimension of equal opportunity policies. In addition, we include one indicator that is an approximate measure of aggregate labour market performance in terms of employment and which serves in turn as a proxy of social integration. We start our demonstration with labour market performance measures, as depicted in Figure 9. Each performance goal coincides with a theoretical benchmark that is either 1 or 0 depending on whether the chosen indicator is expressed as a 'positive' target (to be achieved) or as a 'negative' benchmark (i.e. an output/outcome indicator to be minimized). In the case of the male share in part-time work, equal part-time shares for women and men are considered to be the goal, and so the theoretical benchmark is expressed as half of total part-time work.

Figure 9: Performance Dimensions and Indicators for the Radar Chart Approach (Var 1)

Performance dimension	Performance Goal	Performance indicators and theoretical benchmark (BM in parentheses)
Overall labour market performance / social integration	maximization of employment	• employment as percentage of the working-age population in the core age group 25-54 years (BM =1)
Youth integration into Labour Market System	maximization of young people's (sustainable) integration into labour markets or the education system	• unemployment of the age group 15-19, as percentage of respective working age population (BM= 0) • unemployment of the age group 20-24, as percentage of respective working-age population (BM= 0)
Gender equality in the labour market	minimization of gender inequality in employment systems	• female employment as percentage of male employment in the age group 25-54 (BM= 1) • male share in part-time employment (BM= ½)
Duration of unemployment	minimization of long-term unemployment	• long-term unemployed as percent of the labour force in the age group 25-64 (BM= 0) • long-term unemployment as percentage of all unemployed in the age group 25-64 (BM = 0)

The male part-time indicator is undoubtedly the most controversial of this set of indicators. It is clear that quantitative indicators of part-time work do not address quality issues (e.g. job security, wages, training etc.). It is also clear that the selected indicator does not tell us anything about the gender distribution of 'good' and 'bad' part-time jobs. We are also aware that it is a rather risky assumption that men should bear half of the burden of part-time jobs, a high proportion of which is in fact precarious and low-paid. It would have been preferable to work with an indicator that addresses improvements in women's labour-market position, such as the share of women in full-time work and the gender pay-gap. However, data limitations forced us to use the selected indicators as a crude proxy, their limitations notwithstanding.

The two indicators of youth unemployment might also be criticised for not reflecting educational enrolments (which could be expressed as a positive benchmark). This criticism is mitigated to some extent by the disaggregation into two age groups. Another, more serious objection concerns the indicators of long-term unemployment and gender (in)equality. In each case, there is high interdependency between the pair of indicators, which prevents their use in regression models. It would of course be preferable to use independent indicators, but this will have to wait until radar-chart models are more highly developed.

Finally, some observations on the issue of weighting would be appropriate. The first objection is that there is no justification for giving each of the performance dimensions the same weight, since the relevance of these dimensions remains theoretically, empirically and politically controversial, the diversity of national 'problem mixes' and preferences prohibiting the use of universal performance scales. The second objection is that, even if the sub-indicators could be weighted equally, there is no theoretical argument for giving unemployment a higher total weight than employment; our demonstration model, however, selects three sub-indicators relating to employment and four sub-indicators relating to unemployment.

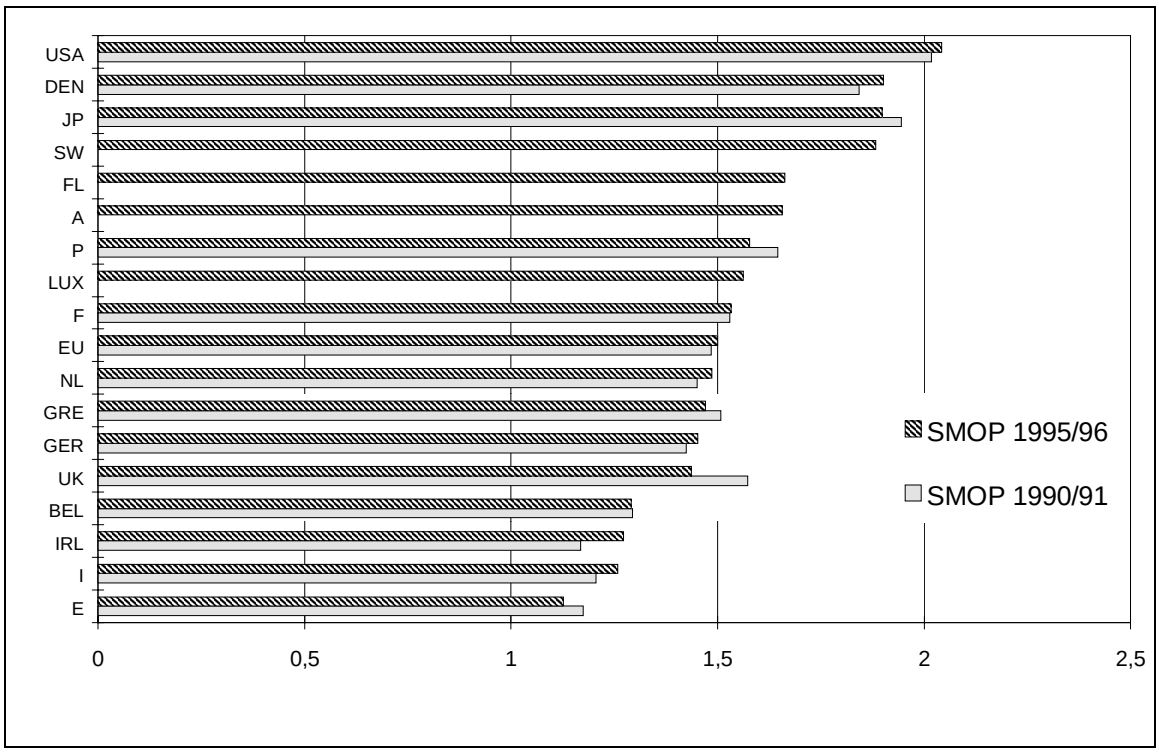
These theoretical objections are justified in principle, although they can also be applied to virtually any kind of fairly broadly based performance assessment. Furthermore, with regard to our example, long-term unemployment can be interpreted as an important structural component of employment regimes. In other words, it is also very informative with regard to overall labour market performance. This feature mitigates somewhat the uneven weighting of employment and unemployment. However, there seems little hope of reaching a consensus on the theoretical justifications for the relative weighting of highly-aggregated indicators such as employment or unemployment rates. The somewhat arbitrary element in our SMOP weighting seems, nonetheless, justified, since it is made explicit and produces the uniform standard required for the purposes of performance comparison. It can also be argued that the SMOP approach is better suited to more homogeneous settings (e.g. across regions within one country) with a set of (decomposed) 'micro-indicators' (e.g. the number of job losses, vacancies filled etc.). The SMOP approach could also conceivably be developed further by using some of the techniques of principal component analysis in order to obtain some clues as to interdependencies and weights that might be useful in the selection of indicators.

5.2.1 Demonstrations of the Radar Chart Approach

The following demonstration uses the set of indicators discussed above. At a later stage, we will examine the sensitivity of the results to variations in the indicators used.

The obvious option of applying the chart method to rankings is displayed in Figure 10, using the results of the SMOP calculated for the EU-15 (or 12), Japan and the USA in the early and mid-1990s. The theoretical maximum value of the synthetic indicator in our example is 2.74, which is clearly not attained by any of the countries considered. The high-performer group is the most interesting. The coexistence of the US, Denmark, Japan and Sweden in this group obviously indicates that absolute superiority does not lie with any particular employment regime, since this group consists of countries representing the liberal market employment regime (US), the state market regime (JP) and the welfare state regime (DK, S) (cf. also Schmid 1998).

Figure 10: Comparison of SMOP for the EU, USA and Japan (SMOP-Var1)



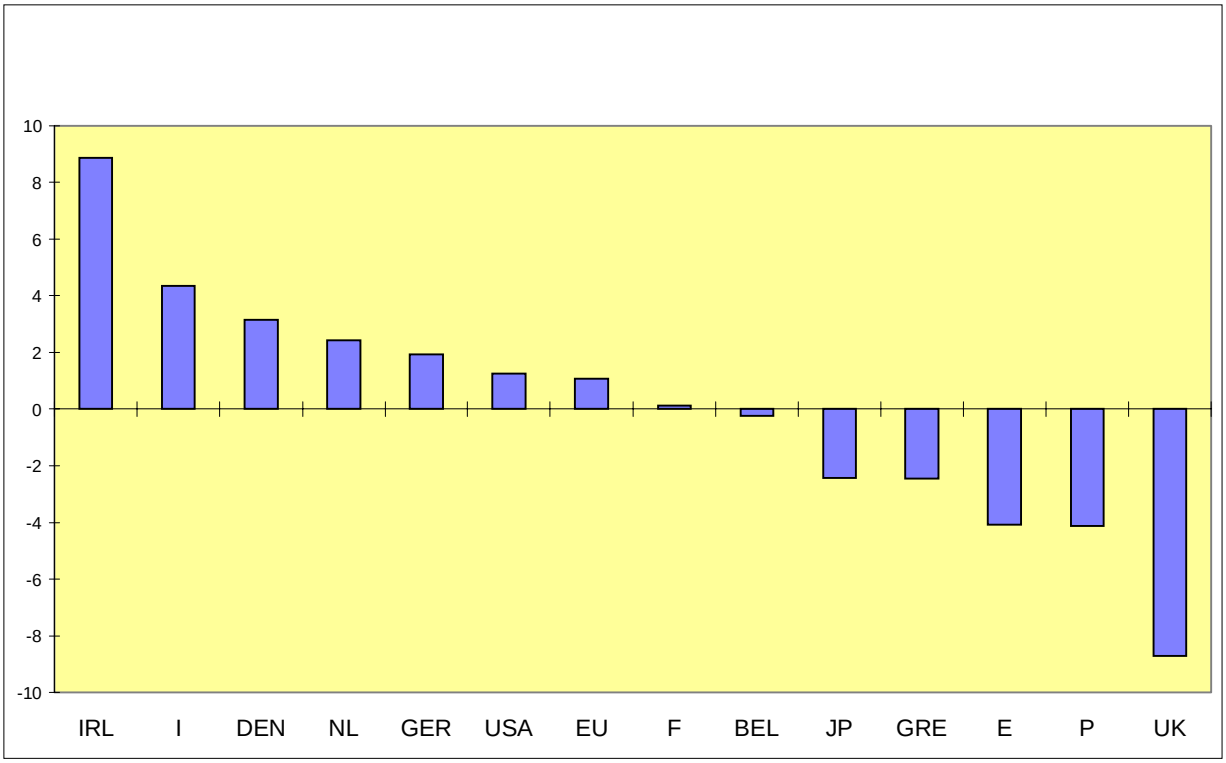
The changes in performance between 1990/91 and 1995/96 are depicted in Figure 11. The most striking results are, firstly, the performance of Ireland as the most improved performer and, secondly, the deterioration in the British performance - probably contrary to many expectations. The reasons for this - in terms of the indicators used - are displayed in Table 1 and will also be discussed below.

Table 1: Raw Data of Radar Chart Indicators (SMOP Var-1)

	TOTEMPR(25-54)		LTUR		LTUSH		YUR95/96		YUR90/91		F/M-EMP		M-PT	
	1990/91	1995/96	1990/91	1995/96	1990/91	1995/96	14-19	20-24	14-19	20-24	1990/91	1995/96	1990/91	1995/96
B	71,7	73,8	4,1	5,4	68,5	68,1	2,3	11,7	1,6	8,0	61,6	70,1	20,1	17,6
DK	84,0	81,7	3,1	2,2	35,3	34,5	5,8	8,6	4,3	12,0	91,9	86,9	28,5	34,1
GER	73,6	76,9	2,1	4,2	49,1	51,7	2,4	6,3	1,0	2,8	68,6	75,9	10,3	13,7
FIN	86,9	75,1	n.a.	4,2	n.a.	43,7	17,4	23,6	5,6	9,2	93,6	94,0	32,2	34,9
F	77,4	77,0	3,4	4,6	44,6	45,4	3,1	15,7	3,0	12,2	72,5	77,9	20,2	21,3
GR	68,5	68,8	2,5	3,5	48,6	52,1	5,5	15,7	4,5	14,1	51,6	54,6	38,9	37,5
UK	79,0	77,2	2,4	3,8	32,9	50,3	8,2	11,3	7,3	10,7	76,6	82,0	14,9	14,3
IRL	60,2	64,5	9,3	7,1	65,5	66,8	5,8	11,7	6,9	15,6	48,5	60,4	28,2	26,8
I	64,9	65,2	4,2	5,4	67,5	63,7	7,3	17,5	6,8	17,9	50,1	56,3	29,2	28,5
L	71,8	71,9	0,7	0,7	31,7	26,0	0,3	3,7	n.a.	n.a.	51,7	54,9	13,4	12,7
NL	71,2	75,0	3,3	3,1	53,3	51,6	6,8	8,0	6,0	6,7	54,7	69,5	24,8	22,8
A	79,0	79,9	n.a.	1,3	n.a.	31,2	3,2	4,0	1,5	3,0	75,3	77,7	10,3	13,6
P	77,4	78,3	1,3	3,2	42,8	55,1	4,0	10,0	3,3	6,3	69,9	77,8	26	27,1
S	91,6	82,6	n.a.	1,6	n.a.	22,8	5,7	12,3	3,5	4,9	95,6	96,5	18,9	23,5
E	61,0	59,1	6,4	11,1	51,4	58,7	10,4	24,1	6,8	19,7	43,6	51,3	20,5	24,9
EU	74,5	73,8	3,6	4,1	49,3	48,1	5,9	12,3	4,4	10,2	67,1	72,4	22,4	23,6
JP	79,6	79,3	0,4	0,4	16	19	1,4	4,2	1,2	2,9	65,4	66,3	31,8	30,2
USA	79,7	79,7	0,8	0,7	6	9	9,3	7,0	7,9	8,2	79,2	82,4	31,4	33,7

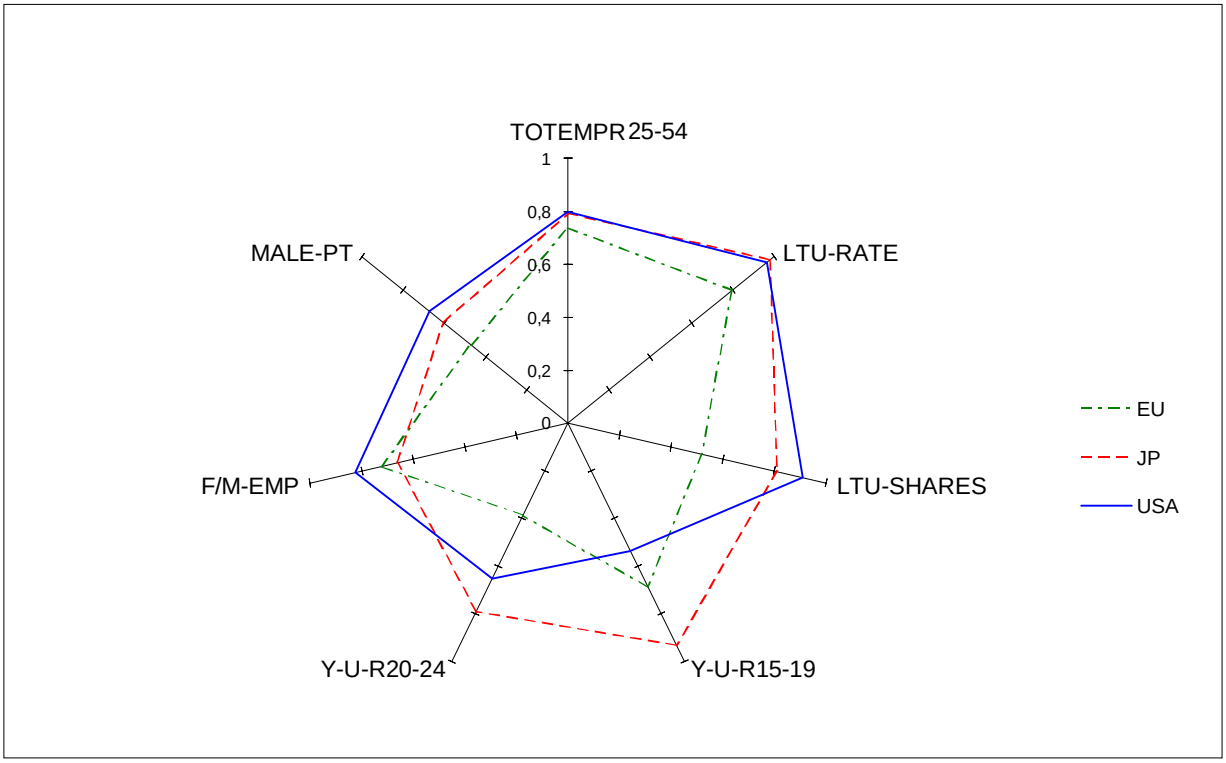
SOURCES: ELFS, Employment Outlook, several years; ILO Yearbook of Labour Statistics 1996; own calculations

Figure 11: Rates of Change in SMOP 1990/91-95/96 (in %) for the EU, USA and Japan



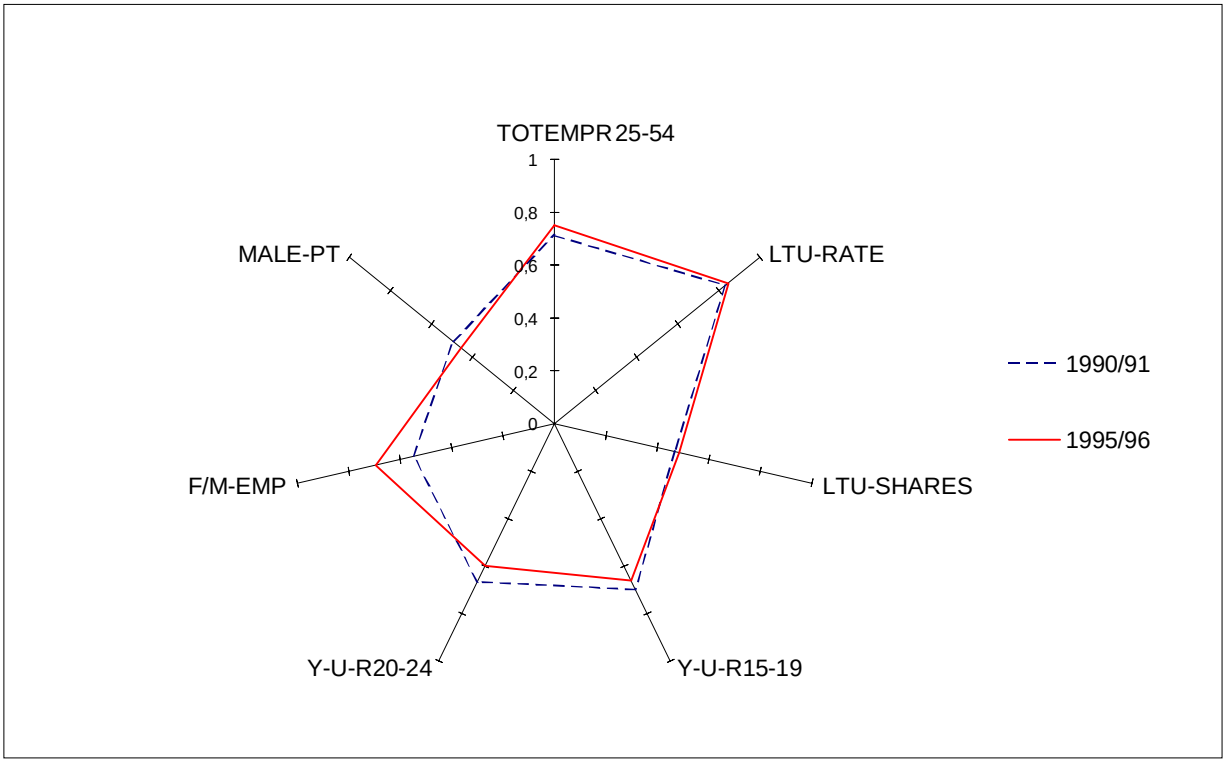
Besides such (numerical) comparisons of a large number of countries, one of the main potentials of the radar chart lies in comparing the performance of two or three countries within one chart. In the following example (Fig. 12), the SMOP of the EU, JP and the US is displayed for 95/96. As can easily be seen, the EU has the smallest overall surface and thus a worse performance than Japan and the US. This is easy to understand, since the US performs relatively better in all dimensions, with the exception of the youth unemployment rate in the teenage group. On the other hand, the better overall performance of JP is due to the European performance gaps in LTU and youth unemployment.

Figure 12: SMOP Comparison of the EU, USA and Japan (1995/96)



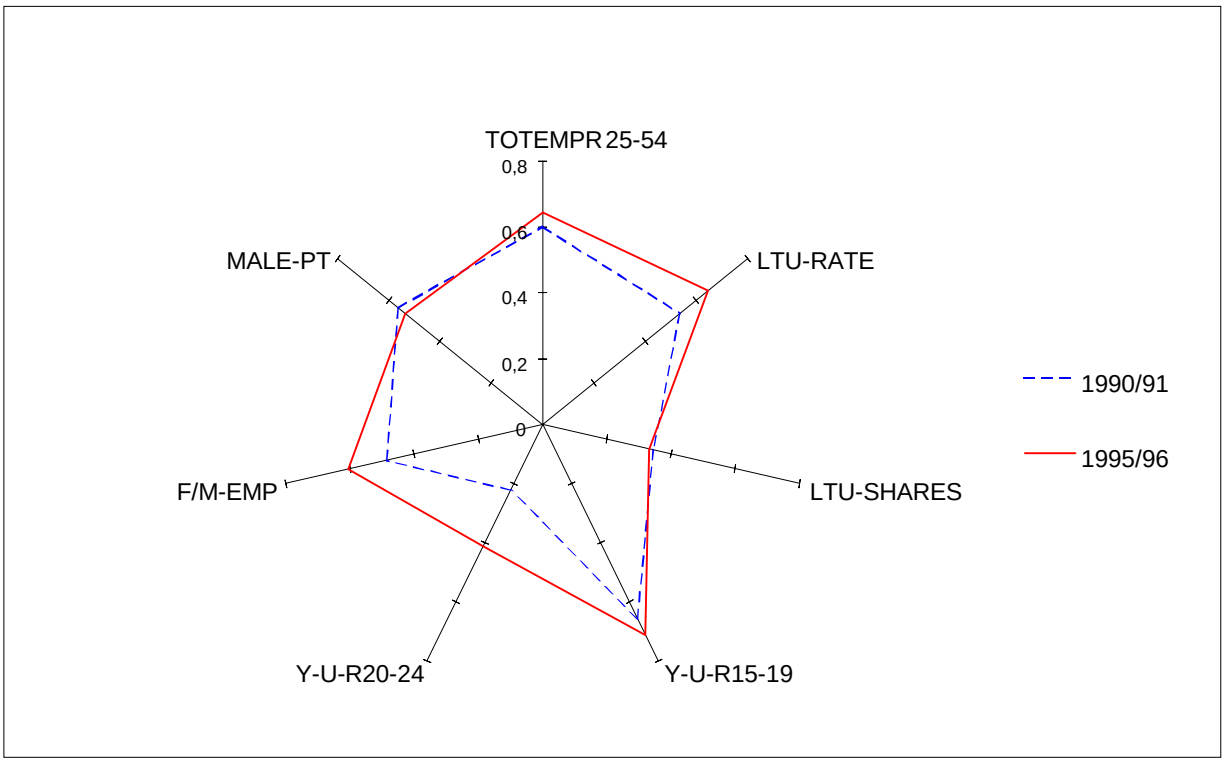
Finally, the option of comparing a country's performance over time will also be briefly illustrated. Let us first consider the Netherlands (Figure 13), which is in a not so favourable middling position in the total ranking (cf. figure 10). What is striking about the Dutch case is the improvement in the female-male employment ratio and the deterioration in youth unemployment. This can be cautiously interpreted as an indication of a trade-off between the promotion of female and youth employment which could be used as a hypothetical starting point for deeper analyses of policy inputs, outputs and impacts.

Figure 13: SMOP Netherlands



Note: The male part-time indicator for the early 90s relates to 1992, since 1990/91 was a break in the statistical series .

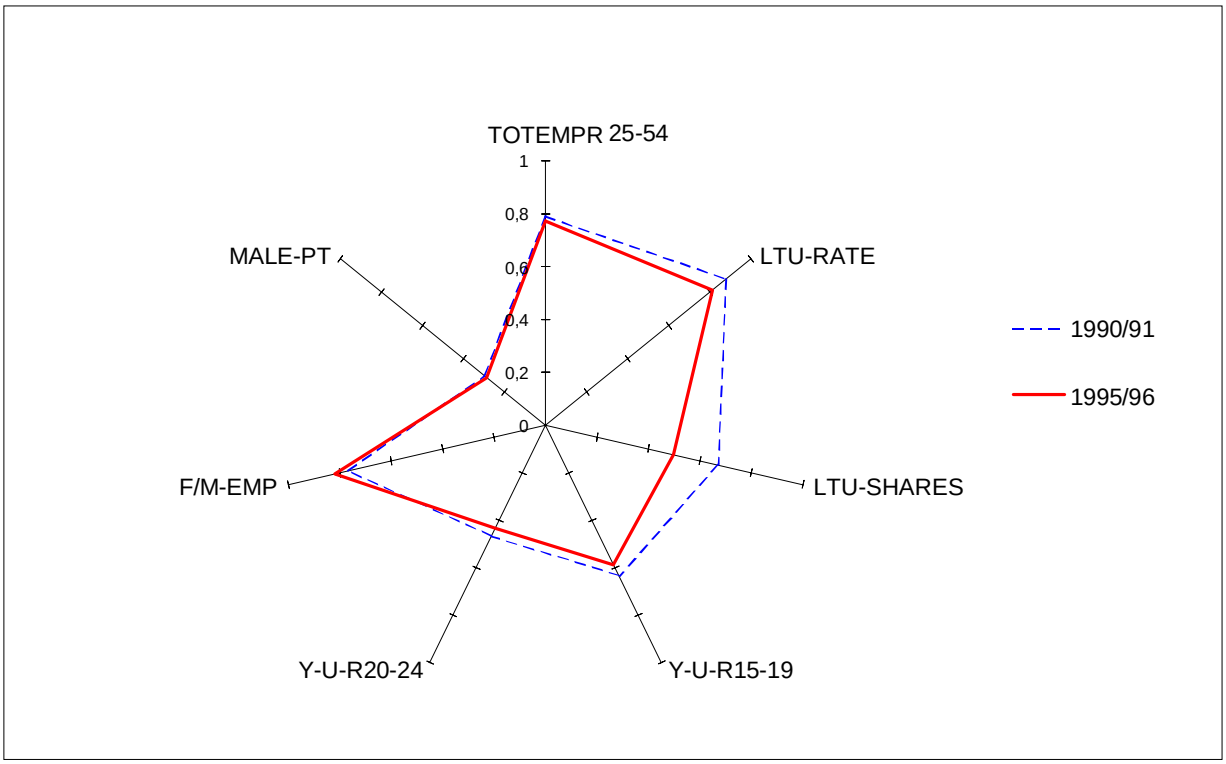
Figure 14: SMOP Ireland



A brief look at Ireland (Figure 14) as the country with the most improved performance between 1990/91 and 95/96 (cf. Fig. 11) would also seem to be appropriate. The relative improvement in 5 out of the 7 performance dimensions (the exceptions are LTU shares and the male share in part-time work) are responsible for this ranking. It is obvious even without looking at the absolute numerical values that the halving of the youth unemployment rate in the 20-24 age group, the high increase in the relative female employment share and the decrease in the LTU rate are the biggest contributors to this result.

The deterioration in the UK's performance (Fig. 15), on the other hand, relates to *all* the indicators used, particularly the LTU dimension, with the sole exception of the female-male employment ratio.

Figure 15: SMOP UK



Let us now look at a final alternative example with some more highly developed indicators that are not available for all the countries previously considered. The following changes will be made.

Firstly, the two gender indicators used previously are replaced by a measure of the gender pay gap and the dissimilarity index. The gender pay gap is expressed as the ratio of average annual earnings of women to those of men, covering all levels of educational attainment, for the age group 25-64 and relates to 1992[47]. The dissimilarity index, in contrast, refers to 1990 since more recent data seem not to be available for the EU countries.[48] The difference in the reference

[47] More recent results (1995) are currently available only for Sweden, Spain, France and the UK, cf. http://europa.eu.int/en/comm/eurostat/compres/en/8397/6308397a.htm. According to this press release, data on the other EU countries will follow.

[48] In the case of the USA, in contrast, fairly good and recent (1995) data on the dissimilarity index are presented

year is justified as the index of dissimilarity (and also the index of segregation) is relatively stable over time and has been changing only very slowly[49]. In the original index, maximum segregation is expressed as 1, so the inverse value of the index is displayed in the chart in order to obtain a positive benchmark.

The second change relates to the indicators of youth integration. The unemployment rate in the teenage group is dropped and replaced by an indicator of educational attainment. This is the OECD indicator of the share of young people successfully completing a full course of secondary education (private or public) at normal age, but related only to graduates of upper secondary education (and not to the 'general' or vocational training/apprenticeship categories which are also included in this OECD statistic).

The other indicators remain unchanged.[50] Given this set of indicators, it was possible to calculate the SMOP for the early 90s for the US and 10 EU countries (excluding Italy, Luxemburg, Austria, Finland and Sweden). The results of this new SMOP are shown in Table 2, which also includes the figures and ranking for Version 1.

Table 2: Performance Comparison of SMOP-Var2 and -Var1 (early 90s)

	SMOP-Var2	SMOP-Var1
USA	1,69 (1)	2,02 (1)
DEN	1,64 (2)	1,84 (2)
GER	1,51 (3)	1,42 (7)
GRE	1,50 (4)	1,51 (5)
F	1,47 (5)	1,53 (4)
UK	1,43 (6)	1,57 (3)
NL	1,35 (7)	1,45 (6)
IRL	1,25 (8)	1,17 (10)
BEL	1,23 (9)	1,29 (8)
E	1,22 (10)	1,17 (9)
Mean	1,43	1,50
Standard deviation	0,17	0,26

Notes: MAX SMOP = 2,74; rank order in parentheses
Sources of new variables: OECD 1995b;
Wooton 1997, Rubery et al 1996.

---

by Wooton (1997) and were used as our data base for the radar chart above.

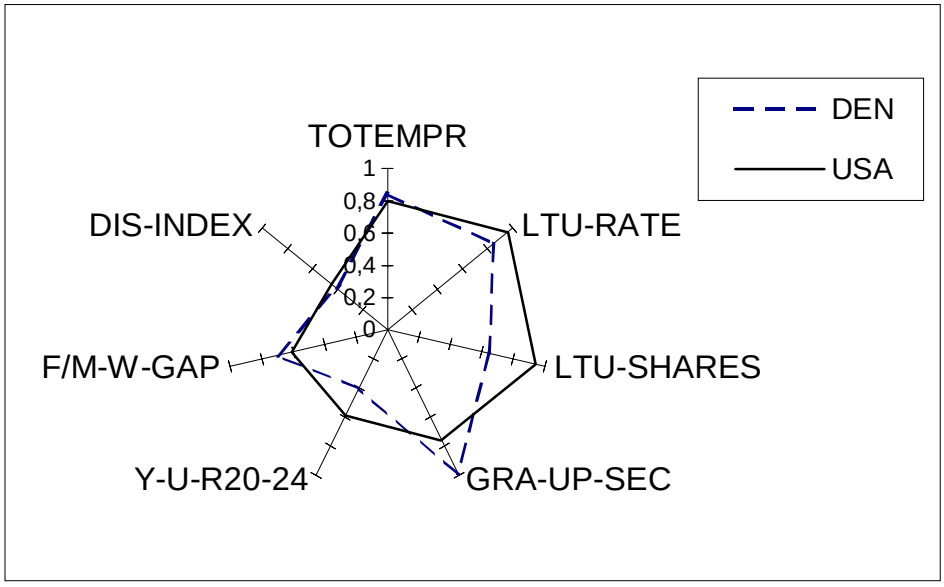
[49] Segregation indices disguise the real dynamics of segregation, cf. Rubery, Fagan and Maier 1996.

[50] There are of course many more conceivable options. For instance, a second indicator of overall employment regime performance could conceivably be included in addition to the total employment rate in order to obtain a more balanced distribution of the performance dimensions included. Possible indicators might be a (relative) poverty indicator (as a measurement of social integration) or labour productivity (as a measurement of economic prosperity).

This comparison shows that the SMOP is sensitive to the choice of indicators, which is hardly a surprise. Firstly, from a comparative perspective, Version 1 produces not only a higher average performance (in terms of the identical mathematical benchmark of both versions) but also greater variability in the results. Secondly, the change in the rank order is in part the result of changes in individual countries' performances caused by the changes in the performance indicators used and in part the result of the relative shifts within the country sample. In the case of Germany, for example, it is less the maximum performance in the education measure and the somewhat better average performance in the equal opportunity dimension than the relative deterioration in the other countries' performances that accounts for the higher ranking. A similar statement holds true for the UK, which performs better in the new version with respect to youth integration but worse in the gender indicators.

So what can be concluded from this short consideration? In simple terms, it is clear that the better the selection and operationalization of the indicators, the better the radar chart measure will be! From our point of view, this seems to be more crucial than problems of weighting, which will always remain controversial. To end this section, the radar charts (new version) for the two best performers, Denmark and the USA, are reproduced below simply for demonstration purposes (Figure 16).

Figure 16: SMOP (Var-2) for Denmark and the USA (early 90s)



5.2.2 The Application of the Radar Chart Approach to Public Employment Services (PES)

Another attractive application of the SMOP-approach is the performance analysis of public (and private) employment services. This can be done either in a cross-national design or across regions within one country. As already noted, the latter will be a more favourable option because of the (relative) institutional homogeneity that shapes the framework conditions of PES activities. In contrast, cross-national performance comparisons of the PES have to take account of different institutional settings; it would also be more difficult to control for economic

conditions. One of the institutional factors, for instance, is the relative importance of open self-service systems for job search. Probably a more important argument for confining any PES-SMOP approaches to national levels, for the time being at least, is the lack of strictly comparable data due to non-standardized monitoring systems.[51] The ambiguity of indicators measuring (aggregate) market shares is also important, since the relative importance of different recruitment channels varies across economic sectors and branches as well as across nations, making it generally difficult to assess the comparative performance of different market shares (see Walwei 1996; Mosley and Speckesser 1997). Probably the best option, therefore, is to use performance indicators taken directly from PES performance management systems (management by objectives/management by results) which are becoming more and more widespread in OECD countries. The quantified performance target of the management system would then serve as the benchmark or the maximum performance on the radar chart.

In the light of the range of PES functions, the possible practical applications and operationalizations of the SMOP approach to PES would appear to be fairly numerous. Here, however, we will concentrate on the matching function as the PES core activity. In accordance with the main thrust of the New Public Management approach, the SMOP should be an ideal instrument for assessing the 'three E's' of economy, efficiency and effectiveness. We will drop the economy aspect[52] and examine only the latter two. Two brief observations are appropriate at this point. Firstly, a distinction needs to be made between organizational efficiency and the efficiency of labour market policy measures, in particular indicators of cost-effectiveness (see Schmid 1996). The number of placements made is not the only indicator of the effectiveness of placement services; quality issues such as the satisfaction of clients, both employers and the unemployed or jobseekers, also need to be taken into account.

The following example is a hypothetical one, i.e. it does not relate to any real performance (though it uses some standard measures), expressed as percentages of the (hypothetical) benchmark.

The following measures of efficiency will be included:

- the number of jobseekers per PES employee (JOBSEEK/E) indicates the work load, or, in positive terms, the availability of competent resources (Schmid 1996)[53] and is also an important control variable;
- placements per net costs of placement - in the absence of any highly developed cost accounting methods (to be applied in a comprehensive controlling system), the proxy denominator for costs will be 'per employee' (PLACEM/E); it should be noted, however, that if efficiency is attributed a higher weight (or importance) than effectiveness, and if the indicator is operationalized as an aggregate measure (regardless of different target groups), it

---

[51] That is not to say that progress is not being made. In the EU considerable steps have been taken to improve the comparability of labour market policy monitoring systems.

[52] The distinction between economy and efficiency is sometimes blurred, so a brief note on the difference is in order at this point. "Economy" concerns the the provision of resources for the goals laid down; an economic organisation provides these factors in sufficient quantity and quality at least cost. In other words, economy is a notion used to measure the extent to which input costs are minimized. Efficiency, in contrast, concerns the relationship of output to input.

[53] As a rule of thumb, a ratio of 100 jobseekers per placement officer is considered to be adequate in terms of competent ressources, which is however almost nowhere achieved (Walwei 1995: 12).

will favour creaming effects;
- vacancy registrations (absolute figures or registration rate = inflow of registered vacancies/inflow of total vacancies) per costs (employee) (VACREG/E)

Needless to say, these efficiency measures require a comparative standard, either over time or across comparable units, before a useful SMOP benchmark can be established. There is also no question that alternative indicators can be constructed and applied.[54] Moreover, a more comprehensive measurement of PES efficiency would have to address the net effects of PES activities on the micro as well as the macro level (Walwei 1995).

In terms of effectiveness, the following indicators will be used:

- vacancy registrations (VACREG) (absolute figures or registration rates), the "most important 'capital' for PES activities" (Walwei 1996), to be expressed as % of the benchmark;
- placements (absolute figures) or user success rate (placed job seekers/inflow of registered job seekers) (PLACEM)[55] [minus the subsequent indicator];
- the number (or share) of placements of long-term unemployed and of those at a very high risk of becoming long-term unemployed (LTUPLAC); the latter to be derived either on pragmatic grounds (for instance, by screening the personal cumulation of selected risks and/or using a defined completed duration to intensify placement activities) or from statistical models when available (cf. Eberts and O'Leary 1997, on profiling and screening methods).

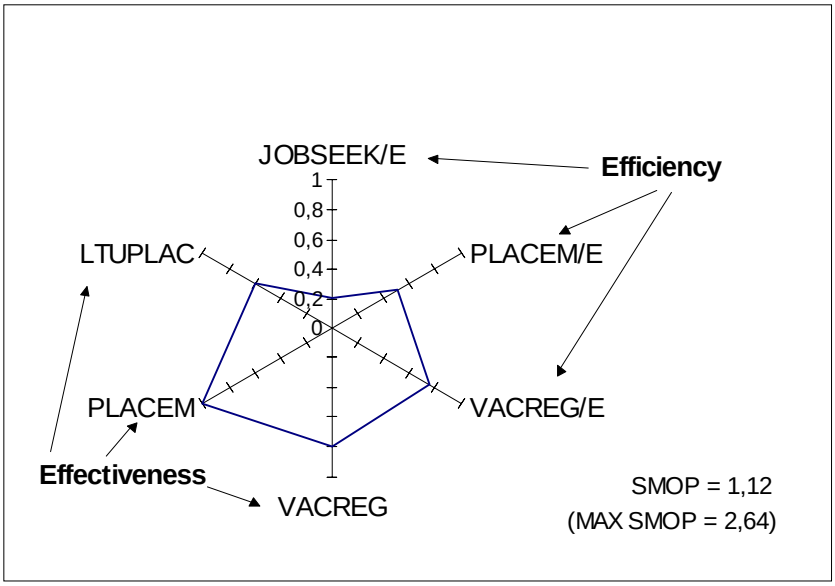
To sum up, even this radar chart, which has only six performance indicators, would provide a complex measure of overall PES performance in terms of effectiveness and efficiency.

---

[54] A very good efficiency indicator would be, for instance, the net inflow of vacancies acquired through pro-active contacts with employers, per costs or per employee. Such an indicator would relate to PES behaviour and not, as is the case with standard indicators such as registration or user rates, to the behaviour of employers and jobseekers. However, the separation of total inflows and this net inflow is difficult to measure. Apart from that, the number of pro-active contacts with employers can of course be an explicit performance target for the PES.

[55] A composite indicator of placements that includes 'regular' placements in the core age group (male, female) as well as placements of target groups, e.g. the long-term-unemployed, the very hard-to-place, the young and the elderly would be conceivable and probably advantageous. Such a composite indicator would have the advantage to demand only one axis of the radar chart (the concrete benchmark would depend on the relative weights of the components included). In contrast, by using several placement rates as single indicators the radar chart loses much of its appeal in terms of parsimony. Still another issue of placements concerns the quality of new jobs. If possible, quality features should be incorporated or alternatively used, such as the share of permanent jobs, or if placement corresponds to qualification level of the job seeker and does not mean a qualificatory downgrading etc.

Figure 17: Employment-Office SMOP (Hypothetical Example)



The demonstration chart would be interpreted as follows. It depicts a fairly good performance in terms of effectiveness, but only an average overall performance in terms of efficiency: The work-load indicator is (realistically) low and the placement-cost ratio indicates that only half of the target performance has been achieved; nevertheless, the efficiency of vacancy registrations reaches three quarters of the benchmark performance. In contrast, the effectiveness measure of overall placements reaches the maximum value, the number of vacancy registrations is satisfactory (assuming that a challenging benchmark is set), and only with respect to the placement of the long-term unemployed would our hypothetical employment agency need to find better ways of reaching the target.

6. Summary and Conclusions

Three core issues have been addressed in this paper. First and foremost, we asked how benchmarking should be conceived and defined in the context of employment and labour market performance and policies, thereby taking into account the conceptual thrust of private sector benchmarking. Second, we addressed the growing relevance of benchmarking in the context of the European Employment Strategy and considered more closely the fields of long-term unemployment, youth integration and the gender aspect of equal opportunities. Third, we suggested the radar chart approach as a tool that could be used to good advantage in benchmarking approaches.

With respect to the first research topic, we suggested that benchmarking could be understood as an approach based on two interconnected pillars, dubbed benchmarking as analytical inventory and benchmarking as an instrument of policy improvement. Regardless of whether results, processes or standards are the object of benchmarking, any benchmarking exercise would remain incomplete if it did not translate new analytical insights into concrete measures to close performance gaps or deficiencies that had been identified. Thus analysis, learning and action are highly interdependent in benchmarking; policy learning is the ultimate and essential goal. Furthermore, we argued in favour of the employment systems and transitional labour market approach, which together provide a complex and timely analytical-normative framework for benchmarking labour market performance and policies. One important implication of these concepts is the emphasis on relational and dynamic indicators. However, irrespective of the theoretical framework applied, benchmarking has necessarily to rely on a wide range of data and methods for its methodological foundation.

As far as the second issue is concerned, we emphasized that account should be taken of the interdependency of the various policies that shape the institutional framework of these fields and determine policy outputs and impacts. This reveals not only the ambitious nature of but also the difficulties inherent in any attempt to benchmark labour market policy. Furthermore, having discussed various theoretical aspects of measurement, we suggested small sets of desirable and available performance indicators that could serve as a starting-point for benchmarking. It is puzzling in fact that labour market policy research still has to contend with data problems, even for some core indicators.

With respect to the radar chart (SMOP) approach, the following advantageous features should be emphasized again:

- the radar chart integrates several performance dimensions into one synthetic (index) indicator;
- it is an excellent means of illustrating or visualizing selected aspects of performance and possible trade-offs between different performance targets;
- the radar chart can be used as a benchmark indicator, demonstrating the performance actually reached against the defined performance objective (i.e. the benchmark);
- the radar chart can be applied for rankings or for performance comparisons across spatial units (countries, regions) and/or over time (change in performance);
- the radar chart approach is suitable for a wide range of specific applications.

We would particularly emphasise the application of the SMOP approach to PES benchmarking, which already seems possible, at least in national contexts, since the required data are available. Moreover, it seems to us a challenge for researchers to look for possible ways of combining the

radar chart approach with the efficiency frontiers approach that was successfully applied to comparison of the performance of the Swedish PES by Althin and Behrenz (1995). Additionally, the further development of the radar chart approach will require and include proper databases, a careful choice of indicators for different purposes and control for context and intervening variables. In sum, the radar chart method presages the important task of developing complex performance measures that both take account of and reduce complexity at the same time.

Finally, what of the future prospects and options for benchmarking? In terms of applicable benchmarking approaches and methods, consideration could be given, for instance, to the institutionalization of quality awards that assess labour market policy organizations and agencies. In the same vein, rating methods already used in some public-sector organizations (see chap. 1.3.3) seem to be a reasonable tool. And it would seem that labour market researchers have barely begun the task of seeking, devising, collecting and adapting the tools and methods required for the benchmarking of labour market performance and policy benchmarking.

The same can be said of the task of defining the actual labour market policies that can most usefully and feasibly be benchmarked. This is also an important task, particularly in light of our argument that (comprehensive) benchmarking will be resource-intensive (i.e. expensive) and must therefore also be evaluated in terms of opportunity costs. Moreover, priorities should also be drawn up on the basis of policy relevance (problem pressure). For instance, public accountability and the balance between citizen and customer focus can be regarded as interesting objects for labour market policy benchmarking, but the analysis of practices that are effective in reducing long-term unemployment might well be considered more important. Additionally, we expect that one important focus of future benchmarking activities in this field will be firms' employment policies, which have already been the object of some attention (Gazier 1998).

However, although the benchmarking of labour market performance and policy is still in its infancy, the European Employment Strategy includes various initiatives that will increase the practical importance of benchmarking and widen its scope for development. There are several potential and challenging topics still to be addressed by benchmarking that are part of the current Employment Guidelines. For instance, the guidelines on the "development of entrepreneurship" and the encouragement of "adaptability in businesses and among their employees" are possible candidates. Moreover, a further development of the Employment Guidelines would be a shift away from input measures (as is currently the case, consider e.g. the targets of employability) to output and even impact measures which would indeed correspond more closely to the conceptual thrust of benchmarking. Finally, closer collaboration and exchanges of views between policy-makers and labour market researchers might also be a very valuable outcome of current and future benchmarking activities.

Bibliography

Albach, H., Moerke, A. 1995: Die Überlegenheit der japanischen Unternehmen im globalen Wettbewerb, Discussion Paper FS IV- 95-4, Wissenschaftszentrum Berlin für Sozialforschung.

Althin, R. and L. Behrenz 1995. En effektivitetsanalys av arbetsförmedlingskontoren i Sverige, EFA: The Swedish Ministry of Labour.

Anxo, D. 1997: Evaluation of two Swedish Wage Subsidy Programs, internal working paper, WZB, Berlin, December 12th, 1996.

Appelbaum, E. and R. Schettkat 1995. Economic Development and Employment in the Industrialized Countries and the Prospects for Full Employment, in: Arestis, P. and M. Marshall (eds.), The Political Economy of Full Employment. Conservatism, Corporatism and Institutional Change, Aldershot: Edward Elgar, pp. 238-255.

Auer, P. and T. Kruppe 1996. Monitoring of Labour Market Policy in the EU Member States. Discussion Papers FS I 96-202, Wissenschaftszentrum Berlin für Sozialforschung.

Beckmann, P. (ed.) 1996. Gender Specific Occupational Segregation, Nürnberg (BeitrAB 188): Institute for Employment Research.

Bellmann, L. 1996. Wage Differentiation and Long-term Unemployment. An International Comparison. Nürnberg, Institute for Employment Research, labour market research topics, No. 19.

Bellmann, L. and Jackman, R. 1996. The Impact of Labour Market Policy on Wages, Employment and Labour Market Mismatch, in: Schmid, G., J. O'Reilly, and K. Schömann (eds.) 1996. International Handbook of Labour Market Policy and Evaluation. Cheltenham: Edward Elgar, pp. 725-746.

Bogan, C.E and M.J. English 1994. Benchmarking for Best-Practices. Winning through Innovative Adaptation. New York and others: McGraw-Hill.

Bovaird, T. 1996. Performance Assessment of Service Quality: Lessons from UK National Initiatives To Influence Local Government, in: H. Hill, H. Klages, E. Löffler, (eds.), Quality, Innovation and Measurement in the Public Sector, Frankfurt a.M., Berlin etc.: Peter Lang, pp. 37-64.

Boxwell, R.J. 1994. Benchmarking for Competitive Advantage. New York and others: McGraw-Hill.

Bryson, J. 1993. A Strategic Planning Process for Public and Non-Profit Organizations, in: J. Bryson (ed.), Strategic Planning for Public Service and Non-Profit Organizations, Oxford etc: Pergamon Press, pp. 11-19.

Büchel, F. 1992. Die Qualität der Wiederbeschäftigung nach ununterbrochener und nach „perforierter" Langzeitarbeitslosigkeit, BeitrAB 162, Nürnberg.

Büchtemann, C. and Brasche U. 1985: Recurrent Unemployment, Longitudinal Evidence for the Federal Republic of Germany, SAMF-Arbeitspapier 1985-3, Berlin

Calmfors, L.and Skedinger, P. 1995. Does Active Labour Market Policy Increase Employment? Theoretical Considerations and Some Empirical Evidence from Sweden, in: Oxford Review of Economics, Vol. 11, No. 1, pp. 91- 109.

Camp, R. C. 1989. Benchmarking - The Search for Industry Best Practices that Lead to Superior Performance. Milwaukee, WI, Quality Press.

Commission 1997a. Joint Employment Report 1997, European Commission, DG V, Brussels.

Commission 1997b. Benchmarking policies for the prevention and reduction of long-term-unemployment, European Commission, DG V, ELC/012/97/EN.

Commission 1997c. Comparative analysis of national performance on integrating young people into the labour market, European Commission, DG V, ELC/009/97/Enrev2.

Commission 1997d. Benchmarking equal opportunities policies, European Commission, DG V, V/D/EVW/mc D(97) 1616.

Commission 1995: The European Employment Strategy: Recent Progress and Prospects for the Future, Brussels, European Commission, COM(95) 465 final.

Commission 1993: Growth, Competitiveness, Employment: The Challenges and Ways Forward in the 21st Century, White Paper, Bulletin of the European Communities, Suppl. 6/93.

Cowper, J. and M. Samuels 1997. Performance Benchmarking in the Public Sector: the UK Experience, in: OECD (ed.) 1997, Benchmarking, Evaluation and Strategic Management in the Public Sector. Papers presented at the 1996 Meeting of the Performance Management Network of the OECD's Public Management Service, Paris: OECD, pp. 11-31.

Cross, R. and A. Iqbal 1995: The Rank Yxerox Experience: Benchmarking Ten Years On, in: Rolstadås, A. (ed.) Benchmarking - Theory and Practice, pp. 3-10.

Dahlberg, L. and C. Isaksson 1997. The Implementation of Benchmarking from a Swedish Perspective, in: OECD (ed.) 1997, Benchmarking, Evaluation and Strategic Management in the Public Sector, pp. 33-48.

Doern, G.B. 1993: The UK Citizen's Charter: Origins and Implementation in Three Agencies, in: Policy and Politics, Vol. 21, No. 1, S. 17-29.

Domptin, C. 1997. Budget: nos voisins au banc d'essai, in: Alternatives Economiques, No. 152, Oct. 1997, pp. 24-27.

Dutch Ministry of Social Affairs and Employment 1996.The Dutch Welfare State from an International and Comparative Perspective. The Hague: Ministry of Social Affairs and Employment.

Dutch Ministry of Economic Affairs 1995. Benchmarking the Netherlands. Test of Dutch Competitiveness. The Hague: Ministry of Economic Affairs.

Eberts, R.W. and C.J. O'Leary 1997. Profiling and Referral to Services of the Long-Term Unemployed: Experiences and Lessons from Several Countries, in inforMISEP, No. 60, pp. 32-39.

Erhel, C., J. Gautié, B. Gazier, S. Morel, 1996. Job Opportunities for the Hard-to-place, in: G. Schmid, J. O'Reilly, K. Schömann, (eds.), pp. 277-307.

Esping-Andersen, Gösta 1994. After the Golden Age: the future of the welfare state in the new global order, Occasional Paper No. 7, World Summit for Social Development , Geneva, November 1994.

EU 1995a, 1996a: Employment in Europe 1995, 1996. Luxemburg: Amt für amtliche Veröffentlichungen.

EU 1997. Chancengleichheit für Frauen und Männer in der Europäischen Union. Jahresbericht 1996, Europäische Kommission, DG V, Luxemburg: Amt für amtliche Veröffentlichungen.

Fagan, C. and J. Rubery 1996. Transitions between Familiy Formation and Paid Employment, in: Schmid, G., J. O'Reilly, and K. Schömann (eds.) 1996. International Handbook of Labour Market Policy and Evaluation, Cheltenham: Edward Elgar, pp. 348-378.

Franz, W. and D. Soskice 1994. The German Apprenticeship System, Discussion Paper FS I 94-302, Wissenschaftszentrum Berlin.

Gaß, G., S. Krömmelbein and A. Schmid 1995. Internationale Maßnahmen zur Bekämpfung der Langzeitarbeitslosigkeit, in: WSI-Mitteilungen, Vol. 48, No. 12, pp. 792-800.

Gazier, B. 1997. The Experience of the „Bilans Sociaux" in French Large Firms: 1977-1997, paper prepared for the Joint Study 'Benchmarking Employment Performance and Labour Market Policies'

by the Employment Observatory RESEARCH-Network (Tronti 1998) DG V of the European Commission.

Gregg, P. and J. Wadsworth 1997. A Year in the Labour Market, in: Employment Audit, Summer 1997, pp18-27.

Hakim, C. 1996. Theoretical and Measurement Issues in the Analysis of Occupational Segregation, in: Beckmann, P (ed.), Gender Specific Occupational Segregation, Nürnberg (BeitrAB 188): Institute for Employment Research, pp. 67-88.

Heise, A. 1995. Wachstum ohne Beschäftigung -Langzeitarbeitslosigkeit eine Folge der „Eurosklerose“?, in: WSI-Mitteilungen, Vol. 48, No. 12, pp. 760-768.

Houseman, S. N. 1995. Job Growth and the Quality of Jobs in the U.S. Economy. in: Labour (IIRA), S93-S124.

Humphries, J. and J. Rubery (eds.) 1995. The Economics of Equal Opportunities, Manchester: Equal Opportunities Commission.

Karlöf, B. and S. Ostblöm 1994. Das Benchmarking-Konzept. Wegweiser zur Spitzenleistung in Qualität und Produktivität, München: Vahlen.

Karr, W. 1997. Die Erfassung der Langzeitarbeitslosigkeit -Ein kaum beachtetes Meßproblem - IAB-Kurzbericht, No.5/97, Nürnberg: Institut für Arbeitsmarkt und Berufsforschung.

Lassnigg, L. et al.. 1994. Möglichkeiten der Evaluation und Effizienzbewertung des Arbeitsmarktservice, Projektbericht des Bundesministeriums für Arbeit und Soziales, Wien.

Leibfried, K. and and C. MacNair 1996. Benchmarking. Von der Konkurrenz lernen, die Konkurrenz überholen, 2nd edition, Freiburg i.Breisgau: Haufe.

Löffler, E. 1996. A Survey on Public Sector Benchmarking Concepts, in: Hill, H., H. Klages and E. Löffler (eds.), Quality, Innovation and Measurement in the Public Sector, Frankfurt a.M. et al.: Lang, pp. 137-159.

Löffler, E. 1995. The Modernization of the Public Sector in an International Comparative Perspective: Concepts and Methods of Awarding and Assessing Quality in the Public Sector in OECD Countries, Speyer: Research Institute for Public Administration (Speyerer Forschungsberichte, No. 151).

Miller, J., A. De Meyer, and Nakane, J. 1992. Benchmarking Global Manufacturing. Understanding International Suppliers, Customers, and Competitors, Homewood, Illinois: Irwin.

Morgan, C., S. Murgatroyd 1994. Total Quality Management in the Public Sector, Buckingham: Open University Press.

Mosley, H. and S. Speckesser 1997. Market Share and Market Segment of Public Employment Services, Discussion Paper FS I-97-208, Wissenschaftszentrum Berlin.

Motohashi, K. and R. Nezu 1997. Why Do Countries Perform Differently?, in: OECD Observer, No. 206, June/July, pp. 19-22.

Naschold, F. 1996. Modernization of the State: Structural Reforms and Innovation Strategies of the Public Sector, in: F. Naschold and C. von Otter, Public Sector Transformation: Rethinking Markets and Hierarchies in Government, Amsterdam and Philadelphia: John Benjamins, pp. 3-91.

Naschold, F. 1995. Ergebnissteuerung, Wettbewerb, Qualitätspolitik. Entwicklungspfade des öffentlichen Sektors in Europa. Berlin: edition sigma.

Naschold, F. and A. Arnkil 1997. Modernization of the Labour Market Organization - Scandinavian and Anglo-Saxon Experiences in an International Benchmarking Perspective, in: J.E. Dolvic, A. Steen (eds.); Making Solidarity Work? - The Norwegian Labour Market in Transition, Oslo: Scandinavian University Press, pp. 279-341.

North, D. C. 1991. Institutions, in: Journal of Economic Perspectives, Vol. 5, No. 1, 97-112.

O'Reilly , J. 1996. Labour Adjustments through Part-time Work, in: Schmid, G., J. O'Reilly, and K. Schömann (eds.) 1996. International Handbook of Labour Market Policy and Evaluation, Cheltenham: Edward Elgar, pp. 566-593.

O'Reilly, J., and S. Bothfeld 1998. Part-time Employment in Germany and the UK -chance or a fate? Comparisons of the Socio-Economic Panel and the British Household Panel. Paper prepared for the 3rd annual meeting of the TRANSLAM working group „Households and Working Time Flexibility, Berlin 30th and 31st of January 1998.

Oakland, J. 1993. Total Quality Management.: the Route to Improving Performance, 2nd edition, Chicester: Wiley.

OECD 1995b. Education at a Glance. OECD-Indicators. Paris: OECD Publications.

OECD 1993a, 1995a, 1996a. Employment Outlook 1993, 1995, 1996. Paris: OECD Publications.

OECD 1993ff. Public Managements Developments. Paris: OECD Publications.

OECD 1992. The Long-Term Unemployed and Measures to Assist Them, Labour Market and Social Policy Occassional Papers No.7, Paris: OECD Publications.

OECD/Céreq 1994. Apprenticeship: Which Way Forward? Paris: OECD Publications.

Oppen, M. 1995. Qualitätsmanagement. Grundverständnisse, Umsetzungsstrategien und ein Erfolgsbericht: die Krankenkassen, Berlin: edition sigma.

Peters, T. and R. H. Waterman 1982. In Search of Excellence. Lessons from America's Best-Run Companies. New York and others: Harper and Row.

Pollitt, C. 1994: The Citizen's Charter: A Preliminary Analysis, in: Public Money and Management, Vol.14, No. 2, pp. 9-13.

Rolle, C. and U. van Suntum 1997. Long-term Unemployment and Social Security: Germany, Austria, Switzerland, USA. Gütersloh: Verlag Bertelsmann Stiftung.

Rolstadås, A. (ed.) 1995. Benchmarking - Theory and Practice. London and others: Chapman and Hall.

Rowthorn, B. and A. Glyn 1990. The diversity of unemployment experience since 1973, in: Marglin, S.A:. and J. Schor (eds.), The Golden Age of Capitalism, Oxford: Oxford University Press, pp. 218-266.

Rubery, J., C. Fagan and F. Maier 1996. Occupational Segregation, Discrimination and Equal Opportunities, in: Schmid, G., J. O'Reilly, and K. Schömann (eds.) 1996. International Handbook of Labour Market Policy and Evaluation, Cheltenham: Edward Elgar, pp. 431-461.

Ryan, P. and C. Büchtemann 1996. The School-to-Work Transition, in: Schmid, G., J. O'Reilly, and K. Schömann (eds.) 1996. International Handbook of Labour Market Policy and Evaluation, Cheltenham: Edward Elgar, pp. 308-347.

Scheuch, E.K. and H. Schimmelpfennig 1974. Skalierungsverfahren in der Sozialforschung, in: König, R. (ed.), Handbuch der empirischen Sozialforschung Bd. 3a, 3rd edition, Stuttgart: Enke and DTV, pp. 97-203.

Schmid, A., S. Krömmelbein, W. Klems and G. Gaß 1994. Neue Wege der Arbeitsmarktpolitik für Langzeitarbeitslose. Berlin: edition sigma.

Schmid, G. 1998. Social Integration by Transitional Labour Markets. Notes on a Theory of Postindustrial Employment Systems, Paper presented at the the 3rd annual meeting of the TRANSLAM network, Berlin 30th and 31st of January 1998.

Schmid, G. 1997. Arbeitslosigkeit und Beschäftigung in Europa: Ansätze zu einer Theorie der Beschäftigungsregimes, in: Statistisches Bundesamt (ed.); Statistische Informationen zum Arbeitsmarkt - Konzepte und Kritik, Anwendung und Auslegung, Stuttgart: Metzler-Poeschel, pp. 15-39.

Schmid, Günther 1996. Process Implementation: Policy Formation and Implementation, in: G. Schmid, J. O'Reilly, K. Schömann, (eds.), pp. 198-231.

Schmid, G. 1995. Is Full Employment Still Possible? Transitional Labour Markets as a New Strategy of Labour Market Policy, in: Economic and Industrial Policy, **16**, pp 429-456.

Schmid, G. (ed.) 1994. Labor Market Institutions in Europe. A Socioeconomic Evaluation of Performance, Armonk, New York 1994: M.E.Sharpe.

Schmid, G. in collaboration with M. Helmer 1998a. The Dutch Employment Miracle? A Comparison of Employment Systems in the Netherlands and Germany, in: L. Delsen and E. de Jong (eds.), The German and Dutch Economies: Who follows Whom?, Heidelberg 1998: Physica, pp. 52-85.

Schmid, G., P. Auer, H. Mosley and K. Schömann (eds.) 1997. Progress in Evaluation Research. Progress in Evaluation Research: Documentation of Two Transfer-Workshops on the „International Handbook of Labour Market Policy and Evalaution", Discussion Paper FS I 97 -203, Wissenschaftszentrum Berlin.

Schmid, G. and K. Schömann, H. Schütz 1997. Evaluierung der Arbeitsmarktpolitik. Ein analytischer Bezugsrahmen am Beispiel des Arbeitsmarktpolitischen Rahmenprogramms in Berlin, WZB-discussion paper FS I 97-204, Wissenschaftszentrum Berlin für Sozialforschung.

Schmid, G., J. O'Reilly, and K. Schömann (eds.) 1996. International Handbook of Labour Market Policy and Evaluation. Cheltenham: Edward Elgar.

Schmid, G., J. O'Reilly, and K. Schömann 1996a. Theory and Methodology of Labour Market Policy and Evaluation: An Introduction, in: Schmid, O'Reilly, and Schömann (eds.) 1996. International Handbook of Labour Market Policy and Evaluation, Cheltenham: Edward Elgar, pp. 1-33.

Schmid, G. and B. Reissert, G. Bruche 1992. Unemployment Insurance and Active Labour Market Policy. Financing Systems in International Comparison, Detroit: Wayne State University Press.

Schmidt, M.G. 1986. The Politics of Labour Market Policy. Structural and Political Determinants of Rates of Unemployment in Industrial Nations., in: Castles, F.G:, F. Lehner and M.G. Schmidt (eds.), Managing Mixed Economies, Berlin and New York: de Gruyter, pp. 4-53.

Schömann, K., T. Kruppe and H. Oschmianski 1998. Arbeitsmarktpolitik und Beschäftigungsdynamik in der Europäischen Union, Discussion Paper FS I 98-203, Wissenschaftszentrum Berlin.

Schömann, K. and T. Kruppe 1996. The Dynamic of Employment in the European Union, in: inforMISEP, no. 55, August, pp. 33-43.

Swift, F.W., T. Gallwey and J.A. Swift 1995. Benchmarking -The Neglected Element in Total Quality Management, in: Rolstadås, A. (ed.) 1995. Benchmarking - Theory and Practice, London and others: Chapman and Hall, pp. 42-50.

Therborn, G. 1986. Why some People are more Unemployed than Others. London: Verso (German Version 1985: Arbeitslosigkeit. Strategien und Politikansätze in den OECD-Ländern. Hamburg: VSA-Verlag.)

Tronti, L. (ed.) 1998. Benchmarking Employment Performance and Labour Market Policies. Final Report of the Employment Observatory RESEARCH Network, Brussels: European Commission DG V.

Trosa, S. 1997. Chairman's Summary, in: OECD (ed.) 1997, Benchmarking, Evaluation and Strategic Management in the Public Sector. Papers presented at the 1996 Meeting of the Performance Management Network of the OECD's Public Management Service, Paris: OECD, pp. 5-9.

Wagner, K. 1998. The German Apprenticeship System after Unification, Discussion Paper FS I 98-301, Wissenschaftszentrum Berlin.

Walwei, U. 1995. Brutto- und Nettoeffekte der Arbeitsvermittlung, Möglichkeiten und Grenzen der Messung ihres Nutzens für Arbeitsuchende, Betriebe und die Volkswirtschaft, in: MittAB, No. 4, pp. 516-26.

Walwei, U. 1996. Performance Evalaution of Public Employment Services (PES), International Labour Office (ILO), Labour Administration Brunch, Document No. 47.

Watson, G. H. 1993. Strategic Benchmarking. How to rate your company's performance against the world's best. New York: Wiley and Sons.

Werner, H. 1997. Die Arbeitsmarktentwicklung in den USA - Lehren für uns? In: MittAB, No. 3, pp.585-600.

Wirth, W. 1991. Coordination of Administrative Controls: Institutional Challenges for Operational Tasks, in: F. Kaufmann (ed.), The Public Sector - Challenge for Coordination and Learning, Berlin/New York: De Gruyter, S. 235-256.

Wooton, B. H. 1997. Gender Differences in Occcupational Employment, in: Monthly Labor Review, No.4, pp.15-24.