

IHS Sociological Series
Working Paper 16
June 1997

Evaluation of Labour Market Policy - Methodological Issues

Lorenz Lassnigg



INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

Impressum

Author(s):

Lorenz Lassnigg

Title:

Evaluation of Labour Market Policy - Methodological Issues

ISSN: Unspecified

**1997 Institut für Höhere Studien - Institute for Advanced Studies
(IHS)**

Josefstädter Straße 39, A-1080 Wien

E-Mail: office@ihs.ac.at

Web: www.ihs.ac.at

All IHS Working Papers are available online:

http://irihs.ihs.ac.at/view/ihs_series/

This paper is available for download without charge at:

<https://irihs.ihs.ac.at/id/eprint/996/>

**Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna**

Reihe Soziologie / Sociological Series

No. 16

**Evaluation of Labour Market Policy -
Methodological Issues**

Lorenz Lassnigg

Evaluation of Labour Market Policy - Methodological Issues

Lorenz Lassnigg

Reihe Soziologie / Sociological Series No. 16

Juni 1997

Institut für Höhere Studien
Stumpergasse 56, A-1060 Wien
Fax: +43/1/597 06 35

Lorenz Lassnigg
Phone: +43/1/599 91-214
e-mail: lassnigg@ihssv.wsr.ac.at

**Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna**

Die **Reihe Soziologie** wird von der Abteilung Soziologie des Instituts für Höhere Studien (IHS) in Wien herausgegeben. Ziel dieser Publikationsreihe ist, abteilungsinterne Arbeitspapiere einer breiteren fachinternen Öffentlichkeit und Diskussion zugänglich zu machen. Die inhaltliche Verantwortung für die veröffentlichten Beiträge liegt bei den AutorInnen. Gastbeiträge werden als solche gekennzeichnet.

Alle Rechte vorbehalten

Abstract

Active labour market policy has gained emphasis in several OECD countries to promote efficient labour markets, and to combat rising unemployment. Consequently the question has gained interest, how to make these policies more effective. During the last ten years several attempts have been made to review the advances in the field of labour market policy evaluation. The paper gives an overview of the field, considering what steps could be taken, and what priorities could be assigned in the development of a rather comprehensive national evaluation practice.

It is argued that each of the different categories of labour market policy evaluation -- policy analysis, impact analysis, and implementation analysis -- should be taken seriously. Priority should be given especially to the development of implementation analysis: setting up sound monitoring systems and performance indicators must be considered as very important steps. Implementation evaluation has a close relationship to impact evaluation because appropriate structures of monitoring and performance indicators are a necessary precondition for impact studies.

Zusammenfassung

Aktive Arbeitsmarktpolitik hat in vielen OECD-Ländern an Bedeutung gewonnen, um die Funktionsweise der Arbeitsmärkte zu verbessern und die steigende Arbeitslosigkeit zu bekämpfen. Daher bekommt auch die Frage wachsende Bedeutung, wie diese Politik effektiver gemacht werden kann. Das Papier gibt einen Überblick über das Feld, und diskutiert welche Prioritätensetzungen und welche Schritte zu einer umfassenden Evaluationspraxis führen können.

Es wird gezeigt, daß jede der verschiedenen Arten von Evaluation -- Politikanalyse, Wirkungsforschung und Implementationsforschung -- erstgenommen werden müssen. Der Entwicklung der Implementationsforschung sollte Priorität gegeben werden, wobei der Aufbau von guten Monitoring Systemen und von Wirkungsindikatoren wichtige Schritte darstellen. Zwischen der Implementationsanalyse und der Wirkungsanalyse besteht ein enger Zusammenhang, da geeignete Monitoring Systeme und Wirkungsindikatoren notwendige Voraussetzungen für das Verständnis der Wirkungen darstellen.

Note:

For helpful comments on earlier drafts of the paper I am indebted to Karl Pichelmann, Patricia Bartholomew, Andreas Wörgötter, and the participants at the OECD-Technical Workshop about Labour Market Policy in the Transition Countries, Vienna 1995.

Contents

Introduction 1

- 1. National Frameworks of Labour Market Policy Evaluation 1**
- 2. The Scope of Labour Market Policy Evaluation 5**
 - a) Areas of Evaluation Activity in General, and Categories of Labour Market Policy Evaluation 6
 - b) Policy Measures and Institutions 11
- 3. Methodological Considerations in the Different Categories of Labour Market Policy Evaluation 14**
 - a) Experimental vs. Non.Experimental Methods - an Unended Quest in Impact Analysis 14
 - b) Micro- and Macro-Evaluation 18
 - c) How to Improve Methodological Standards in Implementation Evaluation? 20
 - d) Policy Evaluation: How to Force the Political System to make Clear Statements about Goals and Objectives? 26

Concluding Remarks 26

References 29

Introduction

Active labour market policy has gained emphasis in several OECD countries to promote efficient labour markets, and to combat rising unemployment. To some extent resources were shifted from passive income support to active measures, so that considerable resources -- on average almost one per cent of GDP -- are spent in this area. Consequently the question has gained interest, how to make these policies more effective.

During the last ten years several attempts have been made to review the advances in the field of labour market policy evaluation. Many issues have become clearer, and the scope of the questions asked has broadened. The following paper will give an overview of the field, considering what steps could be taken, and what priorities could be assigned in the development of a rather comprehensive national evaluation practice. Emphasis for labour market policy as well as evaluation experience varies considerably among countries: some show us well-established practices which had developed during decades, whereas other countries are lacking experience in this field. Consequently, we may take the former as examples for what can be achieved; however, the latter may show us some of the hurdles which must be overcome to develop more sound practices. Therefore, this paper is taken partly from the perspective of a country where rather few exercises in labour market policy evaluation have actually been performed. It tries to outline what can be learned from past experience to apply to further development.

First different types of national frameworks for labour market policy evaluation are outlined, whereby the pros and cons of different frameworks are discussed, and the question is posed which path should be followed in the course of setting up a framework for national evaluation of labour market policy measures. Secondly the scope of labour market policy evaluation is outlined in some detail, arguing that the mainstream of evaluation practice covers a rather narrow scope of the overall field. Different categories of evaluation are discussed with reference to the general evaluation literature as well as to some reviews of labour market policy evaluation. In the third section, the methodology of labour market policy evaluation is considered, focusing on the experimental vs. non-experimental approach, micro vs. macro evaluation, kinds of implementation evaluation, and policy evaluation as the main topics.

1. National frameworks of labour market policy evaluation

As pointed out in the OECD-review of labour market policy evaluation activities from 1989, there have been major differences between countries concerning experience in this field. Some countries

evaluate virtually all important policies (Australia, Canada, U.K., U.S.); other countries -- at least for a decade previous -- did not perform any evaluation (seven countries are listed in this category, including Japan and Italy); and some countries are lying in between the extremes: Germany and Sweden with a more developed evaluation practice; and six countries -- including France, the Netherlands, and also Austria -- are rated to have carried out "evaluations of varying degrees of sophistication on a sporadic basis" (OECD 1989, 49). However, as the policies of the European Union require relatively comprehensive evaluation activities, a process of further development of evaluation practice in the EU member states is going on; e.g., the measures co-financed by the European Structural Funds (ESF) include ex-ante evaluation, process-related monitoring and evaluation, and ex-post evaluation of impact. In addition, the building up of evaluation networks for a more systematic exchange of experience has been started.

In our considerations of methodological issues, we should bear in mind the broad range of practices in different countries, and think about possibilities for improvement in the countries that do not evaluate their policies. One issue of interest in this respect is the different institutional frameworks within which evaluation practice is performed in different countries.¹ This point is a rather important one if a country is planning to set up more comprehensive activities in the area of labour market policy evaluation. Furthermore, it may be said that the institutional setting does also influence some aspects of evaluation methodology.

Two broad "ideal-types" of different institutional frameworks are to be distinguished; with certain countries closely representing these "ideal-types" In the following figure some important characteristics of the US and the Canadian evaluation frameworks are summarized (see figure 2).

Figure 2: Stylized Aspects of Institutional Frameworks for Labour Market Policy Evaluation.

THE UNITED STATES	CANADA
* External evaluation	* internal evaluation
* broad market of evaluators	* definite requirements for evaluation
* various procedures applied	* given set of procedures
* involvement of the academic community	* broad systems generating routine data for monitoring and performance indicators
* experimental methods	* quasi-experimental methods
* publication and evaluation of the results	* not distributed to the public
* "scientific" interests	* feedback to the evaluated institutions
* inputs into the international discussion	

¹ To elaborate on this point, we can refer to the case studies about several countries (US, Canada, Sweden, Germany, Australia, Netherlands, U.K.) which are presented in the OECD 1991 publication.

Putting the emphasis on external, or on internal evaluation procedures, is an important design characteristic of a national evaluation framework. Craig Ridell (1991) has shown some consequences in his comparative analysis of the U.S. and the Canadian evaluation experience. External evaluation means that evaluation activities are performed by personnel outside of the labour market policy institutions. As a prerequisite of this model a professional community of evaluators able to perform the projects must exist, which developed during three decades in the U.S. (Rossi & Wright 1984). This seems to be a rather long time, and to shorten this process in a certain country support from research policy will be necessary. As compared to the development of an internal system, the costs may be more distributed to various budgets in case of external professional development. Another feature of the external-internal alternative concerns the links to the research system, which may easier be built up in external systems. Finally, the issue of an informed public debate seems to be concerned also, because the internal system may rather tend to withhold from the public.

The question, whether evaluations have to be performed by a certain pattern of procedures -- including rules about questions to be answered, timing of projects, methodology to be applied, data to be collected, etc. -- is another important characteristic of the design of an evaluation system. Internal systems seem to develop this kind of regulation rather than external systems, allowing for an establishment of standard procedures for evaluation, but possibly posing some constraints on the further development of methodology.

At this general level, we may discuss some pros and cons of the two different ways of setting up a national evaluation framework. The US -- predominantly external -- model certainly has very much contributed to the improvement of impact analysis accuracy; however, for this model to be realized, a rather large community of evaluators is required that may not easily to be developed in smaller countries; drawbacks of the US model include undeveloped implementation procedures and unsatisfactory feedback mechanisms. The Canadian -- predominantly internal -- model certainly is a good way to develop monitoring mechanisms and performance indicators, and also to easily provide feedback about policy action. However, there may be a tendency of not looking at the full range of problems in the evaluated area, and because of the predominance of internal communication channels also the evaluation of results may be restricted. From these considerations the following conclusion may be drawn: some aspects of both systems may be necessary for the development of a comprehensive evaluation system; perhaps the Australian model may be considered as a combination of the two: it includes a quite comprehensive system of mandatory internal evaluation, with broad monitoring mechanisms, but also gives leeway for external evaluators to conduct projects in this framework.

Another feature of an evaluation framework is the establishment of mandatory requirements for the evaluation of certain policy actions. In several countries evaluation and monitoring is required by

legislation, and this may be a way to accomplish the political will which is necessary for systematic evaluations to be conducted (OECD 1989, 50).²

If we now take a short look at the experience of countries, which have not developed a full evaluation system, we can observe some typical problems in the process of the development of a well-established labour market policy evaluation practice. Björklund (1991) has pointed out that Sweden may be considered a case where the typical problem of a discrepancy between a large amount of money invested in labour market policy measures and a fairly low level of evaluation activities is particularly striking. In assessing the causes for this, beside others, an important point is mentioned: if there are no institutions existing, which collect data to be used as a basis for evaluation studies, it is very difficult to develop evaluation practices. The costs of projects that must collect their own data are relatively high; this holds especially if a country's population is small due to the fact that the relative sample size must also be larger in order to make the necessary breakdowns into subgroups. Therefore the creation of proper mechanisms for collecting and distributing data that can be used by an evaluation community seems to be a very important issue, which may be generalized to other small countries as well.³

In the Netherlands, an improvement of labour market policy evaluation was brought about at the beginning of the eighties. However, a first attempt at a systematic approach within the administration was not very successful, first because the administrative data that were conceived as an information base proved to be unsatisfactory, and second because there was a lack of qualified personnel to carry out the evaluations. Since 1985 a second wave of activities has been under way, and the interest in these activities has been rising because of steps to reorganize labour market policy: delivery of policy measures has been decentralized to various organizations. In this case evaluation has become an important issue for securing quality and efficiency of the policy, and for accountability purposes as well. If we look at the more recent comprehensive review of labour market policy evaluation in the Employment Outlook from 1993, we can see, that this new wave of activities has been successful; evaluation results from the Netherlands are represented in four out of seven categories of labour market policy. However, the problems listed in the Netherlands case study from 1991 may be important for other countries that are trying to set up and develop proper activities in labour market policy evaluation. The main problems were:

- Policy objectives were not specified sufficiently
- Lack of attention to costs of policies
- Focus on direct effects and neglect of indirect effects

² However, some researchers have also taken a critical position in the debates about the value of evaluation, which can be summarized by Wilensky's law: "The more evaluation, the less program development; the more demonstration projects, the less follow-through." (Wilensky 1985, 9; see also Riddell 1991, 49-50)

³ Making the information systems designed for the administration of social security available for programme monitoring and evaluation may be way to overcome this problem. In Austria important steps in this direction have been made during the last decade, with encouraging results; however, these data are rather difficult to handle and are lacking important information (see Pichelmann et al. 1991).

- Lack of attention to core activities such as regular job-finding and training activities
- Too little activity in ex-ante evaluation.⁴

Before trying to draw some conclusions for the design of a national framework, two additional aspects that are important for practical reasons must be mentioned. The first is the time frame for evaluation. It seems to be difficult to coordinate a policy cycle and a time frame of evaluation activities especially when the evaluation community is not very experienced. However there are also more objective trade-offs that can only partly be solved. One is that every practical experience needs time to evolve, however policy very often has to respond to very urgent needs, and therefore no time is left for acquisition of experience. Some general recommendations may be made concerning the problems of time frame: one is that evaluation exercises should be set up at the very beginning of each programme; best through a thorough *ex-ante evaluation* (market exploration, problem analysis, and estimation of desired and possible effects) and development of useful monitoring procedures that ensure a regular flow of information about all important aspects of a programme. Another general recommendation is pointing to continuity, not only in evaluation practice but also in policy development, which mainly means that actual plans should take into account past experience; although this seems to be an obvious statement, it is often not the case.

The second aspect, which is of much practical importance, is the development of personnel resources for the evaluation activities. This question refers to the internal vs. external model. However, in either case, human resources able to perform evaluations have to be developed; and according to the OECD-experience, the attitude in the policy environment seems to be a crucial factor: as pointed out in the 1991 paper "*the key question*" may be rather that "*evaluation gives rise to an adequate and informed public debate*" (OECD 1991, 15). This means that part of a national framework must include the existence of interest in such a debate by the different actors and of an appropriate forum for discussion. In case of Austria some studies show that these preconditions are not met to a high degree.⁵

2. The Scope of Labour Market Policy Evaluation

In this section we will point out that a rather narrow area of the overall scope of labour market policy evaluation, namely impact analysis using quantitative methods, is fairly well developed, whereas we know little about many other areas. This argument is elaborated by looking at different

⁴ See de Koning & de Munnik 1991. p.146, p.154.

⁵ The studies of Zilian & Malle (1993) have analyzed how the different actors on the labour market are dealing with problems of mismatch, or free riding. They show in painstaking details as to how uninformed and contradictory the public debate about these issues is performed.

aspects: (a) a comprehensive taxonomy of different areas of evaluation; (b) distinct categories of labour market policy evaluation; (c) the dimensions of policy actions in institutional terms.

a) Areas of Evaluation Activity in General, and Categories of Labour Market policy evaluation.

If we compare the mainstream of labour market policy evaluation with the more general evaluation literature, we can see a concentration on some specific questions and methods in the mainstream literature.⁶ Whereas in the general evaluation literature a continuous differentiation and broadening of research questions and methodological approaches has taken place, this broadening has not been reflected very much in the labour market policy evaluation field.

In the general evaluation literature, methodological development can be characterized by a twofold structure; the traditional approaches of quantitative impact analysis, and the mainly qualitative approaches of formative evaluation. However, there are not many examples of evaluation activities in labour market policy from the second category.

We may illustrate the broad scope of general evaluation theory by a taxonomy developed by Huey-Tsy Chen (1990), who elaborated the *theory-driven approach* developed by Rossi. Compared to the classical approach of impact analysis, two aspects are brought to the fore: the first is the definition of problems and of goals and objectives of actions. The second is the implementation of these actions. Both were not tackled with in the classical design: goals and objectives were treated as (externally) given, implementation was treated as a "black-box".

Chen distinguishes six broad areas of evaluation which when taken together build a comprehensive strategy of theory-driven evaluation:

1. *Analysis of goals and objectives and their intended results*
2. *Analysis of treatment*
3. *Analysis of the implementation environment*
4. *Impact analysis I: quantitative (outcome oriented)*
5. *Impact analysis II: qualitative (process and implementation oriented)*
6. *Analysis of generalization.*

These areas, which give a good overview about the different kinds of objectives and research questions to be dealt with in the evaluation business, are briefly described in the Table below.

⁶ The OECD (1989) report about the main findings of the evaluation programme had defined evaluation conceptually by impact evaluation. Evaluation "*usually seeks to establish in cost-benefit terms the net effects in both financial and real economic terms.*"(OECD 1989, 44).

Figure 1: Areas of Evaluation Activity (Taxonomy by Huey-Tsy Chen 1990)

1. Analysis of goals and objectives and their intended results

TASK	Clarifying the goals and objectives of a policy. This area refers to the complexity of the goals of many actions; and also to the fact that frequently, various actors are involved in a single activity. This is especially true for labour market policy, where goal conflicts between employers and employees frequently arise, e.g., concerning the equity-efficiency trade-off.	ISSUES	<ul style="list-style-type: none"> * Observation of different views taken upon an activity by different actors (e.g., to observe problems that inhibit cooperation between the actors, or to give an account of fundamental goal conflicts between actors); * to define the intended outcomes or to establish performance indicators; * to find a consensus about priorities; * to analyze the obtainability of certain goals or objectives; * to identify goal conflicts or unintended outcomes.
------	--	--------	--

2. Analysis of treatment

TASK	The analysis of the actions designed to bring about the intended outcome should provide the necessary information to describe the most important aspects of treatment within an action. Any market policy measures -- as, e.g., training or employment subsidies -- can be designed in various manners; moreover, one of the best established findings is that the practical delivery of policy measures tends to vary markedly among local agencies.	ISSUES	<ul style="list-style-type: none"> * Description of the main components of the treatment (e.g., in case of training measures: duration of courses, composition of skill packages, specific targeting and screening procedures, learning sites, training methods, etc); * distinction between conceptual plans for the treatment as opposed to the actually implemented treatment to find out whether the plans have been implemented properly; * information for the design of a satisfactory monitoring system.
------	---	--------	---

3. Analysis of the implementation environment

TASK	All aspects in the environment of an action that may be important with respect to implementation or to outcomes should be under consideration in this area. These dimensions should be also considered to be included into a monitoring system. The elements that link levels of formulation and delivery of an action that are frequently the weakest points of implementation should be taken into account.	ISSUES	<ul style="list-style-type: none"> - Participants (e.g., persons to be served by a measure, target groups, etc.) - Providers (e.g., persons involved in the delivery of a measure: employment service counselors, etc.) - Characteristics of provision (e.g., information channels, selection procedures, etc.) - Organizational frame of provision (e.g., links between participants and providers: mandatory visits at the employment service, entitlements, etc.) - Organizational linkages (e.g., kind of relations between involved organizations: legal basis of relations between employment service, training organizations, and enterprises, etc.) - Micro context of provision (e.g., families of participants, influence of peer-groups, etc.) - Macro context of provision (e.g., regional characteristics, economic cycle, etc.)
------	---	--------	--

Figure 1 continued: Areas of Evaluation Activity (Taxonomy by Huey-Tsy Chen 1990)

4. Impact analysis I: quantitative (outcome oriented)

TASK This is the core area of evaluation research especially in labour market policy evaluation. The main objective is to find out the effects of an evaluated activity. To find out the benefits in financial terms, and to compare the benefits to the costs is the most advanced type of quantitative impact analysis.	ISSUES <ul style="list-style-type: none"> * Meeting <i>internal validity</i> to identify the real net effects of an activity; * to identify the benefits of an activity in accurate terms; * distinction between the intended outcomes and unintended consequences of an action; * the methodological tools of impact analysis require definitions of goals that are frequently not fulfilled.
---	--

5. Impact analysis II: qualitative (process and implementation oriented)

TASK This second area of impact analysis is not included in the traditional "black-box" approaches. The main question here is whether program theory is appropriate: can the intended outcome be reached by the measure as it is conceived?	ISSUES <ul style="list-style-type: none"> * Distinction in program theory between <i>conceptual theory</i> (states the causal relationship that should bring about the efficacy of an intervention) and <i>action theory</i> (states the specific treatment that should put the stated causal relationship into working). * Assessment, whether the action theory has shaped an appropriate treatment on the basis of the conceptual theory. * Empirically observed differences in results may reflect differences in evaluation methods or differences concerning action theory.
---	--

6. Analysis of generalization

TASK In this area the problems of linkage between the field of "research" and the field of "practice" is set up as a distinct area of evaluation activities. Behind this lies the idea that explicit activities of "translation" between the two fields are necessary.	ISSUES <ul style="list-style-type: none"> * All problems of feedback of evaluation results into practice are to be tackled in this area; * to promote understanding each other among evaluators and users; * to secure applicability of evaluation results; * to preclude overgeneralization, or abuse of evaluation results.
--	---

The taxonomy makes clear that the quantitative outcome oriented impact analysis which is the core area of evaluation research especially in labour market policy evaluation should be supplemented by various other tasks and objectives. The methodological tools of impact analysis require the identification of the benefits of an activity in accurate terms. The necessary definitions of goals are frequently not fulfilled. Therefore a close relationship should be considered between outcome oriented impact analysis and goal analysis. An important example which points to this connection concerns the goals and objectives of private-sector employment subsidies: Whereas measures of this kind have been frequently designed for the creation of additional employment, their main achievement may have been rather to improve the relative employment chances for disadvantaged groups (OECD 1988, 56-8). A broad literature has evolved around the topic of dealing with the relation of goals and effects, including the approach of "goal-free" evaluation formulated by Michael Scriven (1972, 1973), which states that the evaluators should choose a set of objective

measures out of the universe of possible outcomes; so it would be possible to analyze a broader set of objectives than those derived directly from the stated goals and objectives.⁷ Another notion concerning the definition of goals is to draw a distinction between "official" and "operative" goals, or between "formal" and "real" ones. Taking up the example of employment subsidies, it is obviously easier to promote publicly the goal of a net employment increase, rather than the more special objective of improving the employment chances of long-term unemployed persons at the expense of short-term unemployed persons.⁸ Some other typical examples which refer to possible tensions between official goal statements and real impacts are: It may not be convenient in political terms to give much public emphasis to the objective of wage moderation; the same holds for the objective of renewal of benefits through the participation in a scheme. In any case, it is widely agreed that the goals and objectives are a critical aspect of impact evaluation.

Qualitative, process- and implementation-oriented impact analysis is another important supplement to the traditional "black-box" approaches. It goes more deeply into the causes why a measure does not produce the intended outcome. If a certain treatment does not work, the reason may be that the action theory is wrong, or conceptual theory is wrong. Empirical studies about impact of certain types of policy measures frequently observe a rather wide range of results, not seldom even pointing in different directions. Consequently the question arises whether the differences reflect differences in evaluation methods or differences concerning action theory. An example which may clarify the question of conceptual theory vs. action theory can be found in the area of private sector employment subsidies, where experience has shown rather severe difficulties in influencing the personnel recruitment patterns of private enterprises; however, it is not clear whether the difficulties are caused by some design features of the programmes, e.g., too low premiums, or whether there are more principal causes at work.

The taxonomy of evaluation activities by Chen works out distinct areas that give a comprehensive overview about the range of different, but interlinked, objectives to be followed, and most of other designs and approaches in the general evaluation literature fit very well into it. We could conceive evolution of the overall evaluation field as a process of "enrichment", meaning that in each of the above sketched areas various activities arise. Possibly the process of evolution of evaluation practices in a certain sector of a certain country may also follow a route which first brings about the involvement of external evaluators, and later on will, through an interactive process, bring back the activity to the internal actors, which then are able to integrate evaluation into their every-day

⁷ The analysis of different effects of active labour market policies given by Calmfors (1994) may be seen as an example for *goal-free evaluation*, because the assessment of a broad range of possible effects is performed on a theoretical background, which is independent from the officially stated goals.

⁸ The possible difference of the officially stated goals to the real goals may be illustrated by an Austrian programme of wage subsidies some years ago: whereas the official goal was "to create jobs", an evaluation of the operative goals showed clearly that the real goal was rather to redistribute unemployment by trying to influence the ranking of long-term unemployed in the queue of job seekers (Lassnigg et al. 1990).

business.⁹ The professional character of services and a decentralized system of delivery of the services may be seen as a precondition for such a development of evaluation, conditions which may be reached also in the field of labour market policy. In any case there should be a closer linkage between evaluation activities in different sectors of society, because learning could take place between them.

The OECD (1991) paper about the state of the art in labour market policy evaluation clearly stated that the traditional preoccupation with the observation of the net-impact of labour market policy programmes is not sufficient for practical purposes. Accordingly three different categories of labour market policy evaluation were distinguished, which can be seen also as a simplified version of the previously cited Chen-taxonomy: (1) policy evaluation, (2) impact evaluation, (3) implementation evaluation.

1. Policy evaluation. The first category, policy evaluation, refers to the analysis and discussion of the broader goals and objectives of labour market policy in the general policy framework of a country in a certain situation. This kind of evaluation has proven to be necessary for at least two reasons. First, careful policy evaluation helps us to avoid too abstract and theoretical considerations, and guides us to a ground of real possibilities; second, policy evaluation helps us to apply the necessary weight to the dimension of the goals and objectives of policies or programmes. An important question is concerning reasonable expectations about the possible outcomes of a policy.¹⁰

2. Impact evaluation. The second category is impact evaluation which is the most developed category in labour market policy evaluation in terms of methodology as well as in terms of accumulated experience. It aims at an accurate observation of the net-impacts of policy measures (which means the actual impact compared to a situation where the measure in question would not have taken place). The impact may be measured by real indicators (e.g. reduction of unemployment, or the placement of long-term-unemployed) or impact may be measured by financial indicators by ideally cost-benefit relations. Despite the most emphasis -- especially in the field of methodology -- has been given to this category of evaluation for several years, it is still not clear whether the most important problems have been adequately solved. The question remains open whether quasi-experimental methods are an adequate alternative to the classical experiment in defining the

⁹ An example where this general line can be observed is the development of evaluation activities in higher education systems. Even when evaluation activities have been started by external actors -- as, e.g., in the U.K. higher education system -- frequently further development brings about a process during which internal mechanisms of evaluation are set up.

¹⁰ As an example, there has been a long-standing discussion concerning the assessment of the strength of the stimulus in US educational programmes aiming at equality of opportunity. On the one hand the definition of what this aim means is difficult; on the other hand, even if we use the experimental approach to assess impact, we had to know previously how strong a stimulus is needed so that impacts can be expected. See for these discussions Mosteller & Moynihan 1972, especially Ch.12 and Rossi, Freeman & Wright 1979, especially Ch.9. LaLonde (1995) also points out that the problem of reasonable expectations about possible impacts of training programmes has not always been taken into account sufficiently.

"control" situation;¹¹ and the other main methodological problems in impact analysis -- the assessment of *deadweight loss*, of *substitution* and *displacement* -- have been seldom analyzed accurately.

3. Implementation evaluation. The third category, implementation evaluation, deals with all aspects of program delivery and development. It is recently seen as the category that needs the most development; despite it is judged recently as being of equal importance as impact evaluation, its state of the art is considered to be by far less developed than the latter (OECD 1991, 13). Implementation evaluation came to the fore, when certain shortcomings of impact evaluation became clear. As impact evaluation was designed principally as "black-box"-model (which compared inputs to outputs, and observed, if there was an effect or not), the results did not tell much, if anything at all, about questions such as the following: why are there little or no effects? Is this due to the principal idea, or only to bad delivery? How to improve a policy measure? etc.

The described distinction between impact evaluation and implementation evaluation is related to a more principle distinction that was made in the evaluation literature between process-oriented or formative evaluation on the one hand and outcome-oriented and summative evaluation on the other. Some authors believe that the methodology to apply would differ deeply concerning these different kinds of evaluation: process evaluation is connected to qualitative methods, whereas outcome-oriented evaluation can hardly be done without quantitative methods; the latter can be performed by external evaluators, whereas in the former involvement of the internal actors and of stakeholders is considered to be essential (cf. Guba & Lincoln 1987).

b) Policy Measures and Institutions

Another aspect where we can identify a narrow perspective in labour market policy evaluation is the dimension of the policy action that is analyzed. Here we can make a distinction between the measures of active labour market policy and the institutional frameworks for bridging demand and supply on the labour market and for the delivery of measures as well (see figure 1).

¹¹ This debate may be represented by some of the meanwhile "classic" articles: Burtless & Orr 1985, Heckman, Hotz & Dabos 1987, Björklund 1988.

Figure 2: Analytical Structure of Labour Market Measures and Institutions

Demand	Bridging	Supply
<u>MEASURES</u>		
Public Sector Job Creation	Adjustment Measures	Early Retirement
Private Sector Wage Incentives	Measures for the Long-term	Reduction of Working Time
Local Incentives	Unemployed	Training and Retraining
	Transition from School to	Training to Remedy Gender
	Work for Youth	Imbalances
		Geographic Mobility
		Assistance
<u>INSTITUTIONS</u>		
	Public Employment Services	
	Decentralized Implementation	
	of Measures	

Source: OECD (1989)

The area of bridging institutions had not often been subject to evaluation during the eighties. In addition "a surprising example of the failure to develop and use rather obvious monitoring data was uncovered by the panel on the Public Employment Service. It found that very few countries regularly collect and monitor job search times and job vacancy filling times, although these constitute prime evidence of the service's effort to improve the functioning of labour markets" (OECD 1989, 50).

However, since 1989 there have been major improvements in this respect. First, we have to mention the pioneering studies about the Canadian and the Swedish employment services (EIC 1989, Delander 1991). Second, during the last few years the OECD has begun to provide comprehensive and comparable descriptions and analyses of the public employment services of member countries (OECD 1993a, 1993b, OECD 1995a). Knowledge about different structures and practices of the public employment service has substantially been widened. Probably the most important general finding of these exercises has been that the provision of active labour market policy measures should be seen in the context of the more traditional activities provided by the public employment service: the links between mechanisms of payment of unemployment insurance benefits and registration procedures for the unemployment register; the practice of notification of vacancies, practices of transition for the long-term unemployed between passive measures and active measures; etc.

Thus, labour market policy increasingly has been perceived as a comprehensive system rather than as a more occasional mix of single measures. This, however, seems to pose more open questions to evaluation practices than are solved by the new perspectives. More fundamentally, we may note that the methodology we have at hand is not very useful for the evaluation of institutions. Rather it has

been developed for the analysis of single measures which are much smaller in scope and, therefore easier to analyze. Some projects, for example the evaluation of the Canadian Employment Service, have tried to evaluate this large scale institution by tackling it as a kind of complex measure. The study that contains several substudies includes various designs and techniques of data collection and analysis, as well as the definition and empirical application of performance measures, and has obtained numerous descriptive and analytical information about the working of the employment service. The generalized results point to rather disappointing issues, one being the observation, that the activity of the public employment service did not contribute to the efficiency of labour market operation, but rather enforces equity objectives, especially job satisfaction by served employees (EIC 1989, 8-11).

Looking at the development of labour market policy we note that the institutional frameworks have come to the fore in several countries; there have also been complaints about the lack of evaluation know-how in this field. In several countries the public employment services have been reformed, or are under a process of reform. An important issue of reforms is decentralization. In that respect the 1989 OECD study has also made rather general points claiming that results about the operation of reforms were not available up to that time.

If we take as an example the Austrian reform, the process of working out the reform plan was accompanied by a rather large scale evaluation project (BMAS 1992). For the first time a comprehensive analysis about the structure of the public employment service was carried out. It included an observation of the goals and objectives both of those imposed by the legal requirements and of those followed by the involved actors (especially the social partners); furthermore, a categorization of the activities carried out by the organization and the measurement of the average resources available for these activities (payment of insurance benefits, services to employers and employees respective to fill vacancies by placement offers, provision of active labour market policy measures, etc.) was performed; and as a consequence of the analyses various recommendations to organizational refinements were made. A closer look at the study and its policy implications reveal some shortcomings that are pointing to more general developmental needs. First, the study was carried on by an external study group originating from another country, a fact which in turn caused inhibitions of the exchange of information between the evaluators and the holders of "indigenous knowledge"¹². Second, the study obtained measures about costs, but did not succeed in obtaining cost-benefit relations. Third, the process oriented observations revealed a great variability of virtually all indicators between the local offices of the employment service but the evaluators were not able to explain this finding. Finally some conclusions -- drawn from international experience -- turned out to be rather neglected by subsequent practice.

¹² The following points are indicating these inhibitions: Available research about the Austrian labour market policy is not fully considered; heavy informal critics in the organization undermine some of the obtained measures.

The recent studies about reorganization of labour market policy in various countries of the European Community constitute an important step in the direction of a more comprehensive evaluation of institutional aspects.¹³ The studies indicate a shift in paradigm towards reform of the institutional framework of policy as opposed to the past practice of adding new measures to the existing ones or shifting emphasis from one measure to another.¹⁴ The Netherlands case study, for example, indicates the development of "active networking" practices between decentralized actors instead of the attempt to centralized regulation and control of all activities. Again, we may infer from this kind of evaluation exercises the importance of a closer look at the various aspects of implementation.

3. Methodological considerations in the different categories of labour market policy evaluation

a) Experimental vs. Non-experimental Methods - an Unended Quest in Impact Analysis

As pointed out previously, in impact analysis a main distinction has been made between experimental and non-experimental methods. In the experimental approach, the assignment of persons either to the experimental group, who are the clients of a measure, or to the control group, who are not clients of the measure, must be performed randomly. In this case observed differences in outcome between the experimental group and the control group ideally can be attributed as effects of the measure -- or, if there are no differences between the groups, the treatment would have proved to have no effect.

In the most widely used non-experimental approach, the treatment group is compared to one or more comparison groups that are matched from existing data bases. This methodology allows comparison of the outcome-variables for the clients of a measure to the same variables for groups that bear similar characteristics. However, the main problem of matching the comparison groups remains; we do not know to what extent the groups are really similar, because of the existence of various unobserved characteristics, which could possibly cause differences in the outcome. The unobserved characteristics may culminate in a "selection bias", a set of variables which influenced the initial enrollment of the clients in the measure, whereas the comparisons were not.¹⁵ This

¹³ See the studies about reorganization of labour market policies in Denmark (Höcker 1994), Germany (Linke 1994), the U.K. (Mosley & Degen 1994), and the Netherlands (Moraal 1994).

¹⁴ The former point is illustrated by the practice in the Netherlands, which was called "*overproduction of policies*" (Moraal 1994, 2); the latter is mentioned by Calmfors (1994, 34) in his discussion of a "balanced portfolio" of labour market policies.

¹⁵ If the random assignment is performed properly, the problem of sample selection bias per definition cannot occur, because members of the treatment group and members of the control group must belong to the same population.

problem is at the core of the current methodological and technical development in the non-experimental approach.

Since evaluation research reached its first peak during the US *war against poverty* in the sixties, there has been an ongoing discussion about the pros and cons of experimental or non-experimental methodology. This debate may be seen as a kind of competition for the most appropriate methodology. Great emphasis was placed on social experimentation, especially since the first large-scale experiment, the New Jersey Income Maintenance Project, had been successfully performed in the beginning of the seventies. Parallel to the period of development of this kind of social experimentation, however, important steps in nonexperimental evaluation occurred too. Probably the most influential book in evaluation research, *Experimental and Quasi-Experimental Designs for Research* by Campbell & Stanley published in 1963, made rather strong points for nonexperimental methods also.¹⁶ During the seventies the bulk of large scale experiments had been run. When their results were evaluated at the beginning of the eighties, heavy criticism arose. Hausman & Wise (1985) documented a comprehensive assessment of these large-scale social experiments, drawing rather discouraging conclusions. Much of the criticism -- apart from numerous weaknesses in design and analysis of the social experiments, was directed at the wide range of questions which should have been answered by single experiments. Hausman & Wise (1985) therefore proposed to refute the attempt of "doubling reality" in future experimental research. They suggested to perform simply designed micro experiments which should test a few treatment alternatives, using large sample sizes and less complicated models of analysis. At the same time when the large-scale experiments were in the stage of field work, attempts were made at developing non-experimental evaluation methods especially setting up databases for a generation of non-random control groups used in the evaluation of the Comprehensive Education and Training Act (CETA). In the middle of the eighties, the experimental design proved to be rather successful in this competition of evaluation approaches, in particular in the public and political arena. There have been several attempts by the advocates of the non-experimental methodology to find techniques for ruling out the problems of sample selection. Nevertheless, despite some encouraging results, the Job Training Longitudinal Survey Research Advisory Panel recommended the experimental methodology for the evaluation of the impact of Job Training Partnership Act in 1985. The OECD (1991) paper clearly stated that experimental methodology is the most reliable way to obtain impact results. Problems of interpretation of non-experimental results, and especially the well-known empirical test of experimental against non-experimental results by LaLonde & Maynard (1987) strongly supported this conclusion.

¹⁶ Shadish & Epstein (1987, 575) showed as a result of a survey in the US evaluation community, that this reference was the only one known by more than 90% of evaluators, more than three thirds reporting, that it had major influenced them.

Reviews of the various studies evaluating the impact of the Comprehensive Education and Training Act (CETA) came to mixed conclusions. These projects were mostly using a non-experimental methodology. For example, Stromsdorfer (1987) showed that there were different results depending on the methodology chosen. The results ranged from significant negative to significant positive effects of training on income of male participants. Additionally, different researchers arrived at different results observed by using the same data to analyze the same problem (Bryant & Rupp 1987; Dickinson, Johnson & West 1987). However, for female participants the different studies showed similar results, namely significant positive effects. LaLonde & Maynard (1987) compared the outcomes of experimental controls to the outcomes of nine different models of matched comparisons for the same measures, using the same data base to evaluate the impact of the National Supported Work Demonstration (NSWD) programme. The results in terms of difference in income between the treatment group and the control viz. comparison groups in case of women, as well as in case of youth are striking: while the control group shows an income which is about 10 per cent higher than in the treatment group, the differences of mean income in the comparison groups range between zero and + 40 per cent compared to the treatment group. In the case of women, one out of six versions of comparison groups shows the same result as the experimental group; in three versions the difference is lower; in two versions it is higher. In case of youth the differences are even larger. As Stromsdorfer (1987, 389) points out, the described differences are due (1) to the construction of the comparison group, and (2) to the specification of statistical models.

Not surprisingly, the advocates of non-experimental methods have since claimed that their techniques are open for substantial improvement, and have criticized the above mentioned comparisons heavily: "*All that LaLonde (1986), Fraker and Maynard (1984,1987) and LaLonde and Maynard (this issue) have shown is that selection bias is an important phenomenon in nonexperimental evaluations. The wide range of estimates produced from their studies simply illustrate that false models produce poor estimates and that different false models produce different false estimates.*" (Heckman, Hotz & Dabos 1987, 424) Heckman and his colleagues argued that the nonexperimental methodology could be substantially improved along two different routes, (1) the use of statistical and econometric techniques to rule out inappropriate models, and (2) the collection of better data for empirical analysis of the selection process. Moreover, an evaluation of the Job Training Partnership Act (JTPA) would allow for important moves especially along the second route. The latter argument points to the close connection between impact evaluation and implementation, at least in the nonexperimental methodology: The more is known about the actual implementation of selection procedures, the better models of the selection problem can be set up.¹⁷

¹⁷ At this point problems are arising in using administrative monitoring data for evaluation purposes, because this kind of data normally provides rather sparse information about implementation issues. Hofer & Pichelmann 1995 have rather successfully used Austrian data for the evaluation of a package of labour market policy measures aimed at rather severely deprived target groups. The following variables could be used for the modeling of the participation decision: age, gender, nationality, problem status, characteristics describing the career previous to the participation (as number of unemployment spells, time out of labour force, etc.).

The ongoing discussion has brought about several analyses and reviews which have considerably refined the arguments.¹⁸ Burtless (1995), in a recent discussion, stresses the point of credibility of results for policy makers and the broader public as the most convincing point for an experimental design. Compared to an experiment nonexperimental methods would include rather strong basic assumptions, which are not so easy understandable by a non-specialist public; e.g., about the "right" equation representing the participation decision, and also about the most useful statistical models. However, the decision about which methodology to apply should take into account pragmatic considerations: "When the direct benefits from improved knowledge are easy to predict and measure, analysts can calculate the financial gains from improved decision making and compare them with the additional costs associated with conducting an experiment. An experiment should be undertaken when the value of the improved decision exceeds the extra costs of the experiment (Burtless 1995, 82).

The advocates of nonexperimental evaluation have indeed developed important additional arguments to underscore their point of view. Using information from the JTPA evaluation, Heckman & Smith (1995) make some points which show remarkable weaknesses of the applied experimental methodology. In addition to the long-standing argument that black-box experiments about effects of treatment provide only limited information about the policy in question, they also argue that precise results are confined to mean outcomes, but cannot account for variability between persons. Moreover they provide empirical evidence for the existence of both, *randomization bias* (the influence on participation by the randomization procedures), as well as *substitution bias* (participants in the control group gain access to similar treatments as the experimental group); and they show severe institutional constraints on proper design of the experiment (e.g., difficulties to find participants in the experiment at random, or programme administrators refusing to apply randomization properly). Overall, these limitations lead to the conclusion that the experiment cannot provide proper estimations of treatment outcomes compared to a situation without treatment, as it is promised by the experimental methodology.

This discussion will continue and we may be curious about the future technical improvements of the nonexperimental methodology. The 1991 OECD report stated clearly, that ethical obstacles -- especially the point that a specific treatment has to be refused to persons who would need it -- preclude the use of random experiments in many countries, "*in fact, the technique is rarely used outside of the United States*" (OECD 1991, 12). However, in many countries nonexperimental methodology is utilized.

¹⁸ See as a review Riddell 1991, or more recently the discussion in the spring 1995 issue of The Journal of Economic Perspectives.

b) Micro- and Macro-Evaluation.

In the course of social experimentation development the scope of interest has apparently shifted from the original emphasis on the generation of knowledge about broad behavioural aspects of the population (e.g. the attempts to test labour supply functions in the Negative Income Tax Experiments) to the mere observation of the impact of specific treatments. This shift makes it easier to conduct experiments; however, it also limits the generation of new knowledge about behavioural aspects underlying the provision of policy measures. Heckman & Smith (1995, p.86) suggest that the limitation of questions to *what works?* vs. *what does not work?*, reflects a more general withdrawal from contributions to the cumulative body of knowledge which would also include understanding *why* what works, works or does not. The contributions and discussions in the Hausman & Wise (1985) volume already referred to this question about the contribution of the social experiments as opposed to knowledge from non-experimental studies, and their views were rather pessimistic.

In any case, the more general questions in labour market policy evaluation which transcend the immediate concern about the functioning of programmes, have seldom been treated within the experimental studies. Moreover, most past evaluations are confined to the micro level of results for participants. Two somewhat different, but conceptually linked courses of study which go further than the questions at the micro level may be distinguished. One deals with the more immediate effects of policy measures on non-participants, especially substitution and displacement mechanisms; the second is more concerned with overall macroeconomic effects of active labour market policy.

The latter has recently gained interest; however, the existing analyses unanimously concede that there is much left to do in this field. In a recent review, it is stated that "*a full assessment of the macroeconomic impact of active labour market policies is hindered by a lack of data and the difficulty in establishing robust causal links between the policies and labour market outcomes*" (OECD 1993, ix). Calmfors (1994) has worked out a comprehensive account of different effects of active labour market policies.¹⁹ Macroeconomic effects may arise by an increase of the efficiency of matching between demand and supply on the labour market or by reducing wage pressures by bridging the gap between "insiders" and "outsiders". Furthermore, the productivity of the work force may be increased by training activities, thus contributing to the competitiveness of an economy and possibly increasing employment and wages. The effective supply of labour may be maintained by reducing the risk of long-term unemployment and a subsequent harm to human resources. However, there is also concern that especially job creation measures and employment

¹⁹ See as discussion of macroeconomic effects also Layard 1986, Layard, Nickell & Jackman 1991, OECD 1993, Jackman 1995.

subsidies may create a mere substitution or displacement of "regular" employment by subsidized employment. Another concern is that there may be only "cosmetic effects" of reclassifying unemployed persons as scheme participants; a mechanism possibly reinforced by adverse incentive effects of participation in an active labour market policy measure.

A closer discussion of methodological problems in this area would exceed the scope of this paper. Yet, a general problem in this respect is the question of causality. The provision of active measures, and consequently changes in expenditure can *reflect* changes in labour market performance as well as it can *influence* labour market performance. Problems of substitution and displacement have also been analyzed to some extent in micro-studies. Mostly the evidence is based on more or less direct questions posed to enterprises, regarding how the availability of measures would have influenced their activities. Studies which report results on these aspects indicate that substitution and displacement seem to be rather high.²⁰

If we look at the recent attempts to assess macroeconomic aspects of the problem, we may note that these analyses are establishing closer links between questions of impact evaluation and questions of the other categories of labour market policy evaluation, namely to policy evaluation. When the overall employment effects of an active labour market policy (or effects on wage formation) are evaluated, the underlying aims and objectives are directly involved. Consequently, the findings may serve as important inputs in the discussion about reasonable objectives. The discussion of the insurance function of active labour market policy measures by Jackman (1995) provides an interesting example in that respect. He states that in addition to their intended direct effects, the programmes of labour market policy could also be used to deal with the problems of moral hazard in unemployment benefits. If participation in a measure after a period of unemployment were required for entitlement to further benefit, this may influence claimants who are able to find a job to prefer this to the participation in a programme -- or as Jackman puts it, the required participation in a scheme "*can replace administrative selection by self-selection ... The disincentive effects of the benefit system would be countered, while the insurance objective would be safeguarded*" (Jackman 1995, p.12).

Moreover, the macroeconomic analyses recently have also begun to establish closer links to general questions of implementation evaluation. As an example, we can refer to the problem of the scale of labour market policy measures, either in terms of expenditure or of number of participants: wage equations have been estimated to test the effect of number of programme participants and average cost of participants on wage formation indicating that both variables may have wage moderating impacts (OECD 1993c, 52). Another example is the question of interactive impacts of specific programmes (i.e. training) and functions of the public employment service on overall employment effects, or effects of the amount of long-term unemployment on the impact of labour market policy.

²⁰ See OECD 1993, 63-64, Breen & Halpin 1989, de Koning et al. 1992.

More evidence regarding these aspects may inform us substantially about the most effective implementation of labour market policy measures.

c) How to Improve Methodological Standards in Implementation Evaluation?

Implementation evaluation deals with the various aspects of the delivery of programmes, or in other words, with the working patterns and mechanisms of institutions implementing the programme. In this category methodological paradigms similar to impact evaluation paradigms have not been developed. Perhaps development will not be possible. The range of questions and perspectives, as well as the range of disciplines and methods is very broad, and the scientific paradigm of evaluation is to a large extent ruled out by a professional paradigm of improvement of practices.

The core objective of implementation evaluation is provision of an accurate account of the measures in question. This includes description of all important aspects of treatment, the analysis of the implementation environment, and finally feedback and generalization. In addition, two more general questions may be identified in implementation evaluation, one is whether the planned actions are to be considered appropriate for meeting the goals and objectives. The second is whether the realized actions are in accordance with the planned actions.

The OECD has proposed a distinction of three types of implementation evaluation:

- monitoring,
- performance indicators,
- in-depth implementation analysis.

Monitoring means to set up a descriptive database about the delivery of measures, e.g. parameters of the measure, recruitment and participation, costs, survival or drop out-rates, post programme performance;

performance indicators mean that some of the descriptive data are related to the targets of the measure, giving some information about success;

in-depth implementation analysis means mainly that some kind of qualitative assessment of programme infrastructure is carried out. As there is the problem that using performance indicators as management instrument may cause adverse or non intended effects on the practice, this kind of qualitative analysis is very important.

Apart from the assessment of the impact of measures, the implementation of regular monitoring procedures has formed another root of development of evaluation activities. According to the OECD (1989) report monitoring should primarily provide the necessary information to operational managers in order that they can achieve better management and control. Monitoring activities are

seen as relatively separate from evaluation, the latter being designed mainly to inform major policy decisions. However, the necessary management information corresponds closely to the management style within an organization. The style in traditional administration requires different kinds of information than the management-by-objectives style. Whereas the traditional style is primarily input-related, the latter is -- or at least should be -- output-related. If we suppose that during the last decades a modernization process has taken place in administrative style, this process should have been accompanied by a process of improvement of monitoring procedures.

The case studies about different national evaluation systems indicate that there is a positive correlation between the decentralized provision of policies, mandatory requirements for periodic evaluation exercises, and the set up of sound monitoring systems. Australia, as an example, operates a programme monitoring system (*Programme Administration and Statistical System*) which provides computerized information on the basis of unit record data of programme placements. The information system contains the following information (OECD 1991, 142-3):

- Administrative details (e.g. geographic location, programme category, commencement and termination dates);
- Participant characteristics (e.g. date of birth, gender, educational attainment, duration of unemployment, identified disadvantage);
- Placement details (e.g. employer/ training provider, industry, occupational details);
- Adjunct: Post programme monitoring survey about labour market status three months after programme termination (this adjunct is not operated for all programmes).

Probably in most countries most of these data are collected and registered somewhere, however the question remains, whether this information is used, and after all whether it is organized in a manner that can be used.

There are indeed indications that traditional administrative information systems do actually not provide the data which are necessary for an assessment of the ongoing practices. The Netherlands reported that a first wave of evaluation activities failed, apart from other reasons because of "*defective registration and information systems*" (de Koning & de Munnik 1991, p.152). In Austria, despite the existence of a large and comprehensive information processing system, it is also difficult to use these data for generating aggregate information about a certain programme because the necessary procedures do not exist. The periodic information which is published about labour market policy measures also is not even sufficient to answer the simple evaluative questions which would exceed the most simple descriptive information about how many participants have been enrolled in raw categories of measures. There does not exist for example a meaningful classification of training programmes, despite the fact that these have been the most important activities for more than a decade. We can argue that in many countries much is to be done in the case of the setting up of proper monitoring systems.

If we take into account that the decentralized provision of measures is one of the dominant developments in labour market policy, and that furthermore an improved coordination of the different aspects of policy -- e.g., the regulations concerning unemployment insurance and the design of active measures -- is an important factor for efficiency, then we must conclude that sufficient information about the working of the system, and therefore the development of proper monitoring frameworks will be one of the crucial points for future development.

The next step in implementation evaluation is the definition of *performance indicators*, which link the descriptive monitoring information to the stated goals and objectives. We can once again refer to the Australian evaluation framework, where the prescribed evaluation cycle for new programmes or pilot programmes from the beginning involves the definition of objectives preferably in objectively measurable terms, of which appropriate performance indicators are derived. These indicators are part of a formal evaluation, and the information necessary to obtain the performance indicators is also included in the management information system about the programme. We have recently observed many efforts to develop and to obtain monitoring procedures and performance indicators also on the regular activities of placement and counseling by the employment service. Hence, the above mentioned lack of attention to these activities -- as opposed to the high attention given to the active measures in the past -- may soon be overcome. Change in the administrative style (which is frequently reported changing towards management-by-objectives) in fact requires some sets of performance indicators. Sweden is a country which has a rather long experience in this respect. Since the seventies, a management-information system has been developed which provides monthly information of the following kind for all units of the employment service:

- Number and characteristics of filled vacancies;
- Job-search- and vacancy-filling duration;
- Development of long-term unemployment;
- Number and characteristics of programme participants;
- Duration of participants' waiting for the beginning of a programme;
- Outcome of programmes with respect to labour market status.

Budgetary allocations were allotted according to the indicators, and some years ago some output indicators were utilized to adjust partly the financing of the operating units. Indicators like the following have been used for this purpose:

- Percentage of clients who leave the register for work;
- Duration on the register before leaving the register for work or a scheme;
- Duration of vacancies
- Percentage of vacancies with at least one proposal for filling it.

Obviously the imposition of a new overall system of performance indicators has practical consequences for the institution and is bound to change the working of the system significantly. As

a consequence, attempts in this direction can cause many rumours, and may take a very long time to be set in action.²¹ Although many countries have taken minor steps to management-by-objectives and decentralization reforms, sufficient information to learn from other countries' experience is not available -- and, we perhaps can add, that such kind of information, even if it is available, often may not be used because of other political priorities.

These considerations lead us to our next category, *in-depth implementation analysis*. This kind of analysis received little attention in the OECD (1989) report. One reason may be that the source of information was the work conducted by panels consisting mainly of country officials. Members of the administration likely take the view that planned measures at least in principle are implemented accurately. They certainly know that there are shortcomings, however they will operate on the hypothesis that on average the delivery is in accord to the rules and regulations. However, as pointed out above, only two years later implementation analysis was regarded to be equal in importance to impact analysis: The OECD 1991 report conceded that implementation analysis remained underdeveloped, but virtually no recommendations were made about this topic. The 1993 review of effects of labour market policy measures refers to implementation research in rather implicit than overt terms, despite the fact that the results are pointing strongly to the need to address this issue.

Indeed there seem to be some hurdles to the further development of in-depth implementation analysis which are not easy to overcome. One is that labour market policy evaluation is strongly influenced by economic concepts, whereas the questions of implementation research are linked to other social science concepts and frequently require interdisciplinary approaches. In addition, the field does not have a set of distinct and well-rounded research questions that are awaiting the right answers. Large parts of the work to be done are in fact "applied" and "descriptive" topics, which are labels partly considered as insults in the academic community. Secondly, a kind of "cultural lag" may be observed, which is due to large differences in evaluation experience in different countries: in the beginning of the evaluation practice, the questions of effects of actions were very appealing to policy makers and managers, and this is reinforced by the strong literature about this aspect. However, it may be rather unpleasant to allow evaluators to have a close look into the more intimate processes within institutions, which is inevitable for implementation analysis.²² If we take the US experience, which may be seen as paradigmatic for the development of the overall evaluation business, we can observe that the call for implementation analysis was reinforced at a rather late stage of development, when the shortcomings of mere impact analysis increasingly could not be

²¹ As an example we may refer to a comprehensive report about the Austrian public employment service which presented some data that can be used as a basis for performance indicators. However this information is heavily objected to within the organization, and despite it had been published some years ago it is officially rather due to "inhibition" than taken as a point of departure for further action (cf. BMAS 1992).

²² This view is strongly reinforced by some experience in conducting a project about possibilities for evaluation in the Austrian employment service (cf. Lassnigg et al. 1994).

overlooked. Without effort at counteraction this development may be repeated in other countries. Finally, we should take into account that an active labour market policy is a relatively young course of action, implemented in many countries not before the late sixties, or even later, and that this kind of action has not been undisputed. Therefore it is understandable that the question about impact came first. When in principle impact is positively established, the question about improvement arises.

However, as Richard Nathan (1991) pointed out convincingly, the notion of putting impact evaluation before implementation evaluation may be misleading. He makes very strong points for an evaluation strategy within which implementation analysis should in any case precede impact analysis. This view is based both on analytical concepts and on empirical observations. In analytical terms it is argued that a sharp distinction should be drawn between measures which are newly set up on the one hand, and established ongoing measures on the other. Different kinds of assessment, including different analytical questions and designs correspond to the different kinds of measures, and this should be due mainly to the well-grounded expectation of different behaviours of implementing bodies connected to each of the types of measures. In case of newly established measures, it is expected that implementation will be rather in accord with the schedule, and the main evaluative question is for the impact, which is intended to occur; but because supposedly new ideas of functioning are underlying the new action, it is unknown if it will work. In addition this kind of measure should normally be small in scale, and therefore rather easily manageable. Nathan calls this kind of assessment *demonstration research*, and supports the experimental paradigm as a useful design for this kind of research.²³ However, in case of ongoing programmes which are broad in scale and cover large institutions, two kinds of analytical questions are formulated, the first concerns how the implementing institution deals with the planned measure (impact of the measure on the institution), and the second concerns the impact on participants in the measure (impact on individuals). According to the existing research, about implementation, the first aspect is considered to be the more important one, because of the normally wide variations in implementing the planned measures that normally occur (which are frequently identified by a certain law or regulation, and the respective budget). "*We need to think sequentially. Implementation/institutional research needs to precede the study of programme impact on individuals in evaluation research. We need to know what it is that may or may not cause the individual impacts we now want to measure.*" (Nathan 1991, 25) Consequently, in many cases the second stage, assessment of individual impact, may be omitted.

This view which radically differs from the conventional view in labour market policy evaluation is strongly reinforced by the classical body of implementation research literature which mainly

²³ It should be noted that the terms demonstration research or demonstration studies are not used always in the same way. Some authors use it rather in the meaning of pilot studies, sometimes implying a strategy similar to experiments but with less stringent methodological requirements. According to Nathan most of the well-known experimental studies fall in the category of demonstration research, which is in his view opposed to evaluation research.

originated in the field of political science, lending proof to the point that the planned schedules of programmes mostly are vague images of the real action once the programme is implemented. This means that especially large scale programmes get their real shape not in the process of formulation of the plan, but during the process of implementation which involves a rather large set of interplaying actors.²⁴ As a striking illustration of implementation problems we may look at the indications for the *substitution bias* which Heckman & Smith (1995, 106) reported on basis of their reanalysis of data from the JTPA-Evaluation. In this project which may be seen as a significant experimental study, *"according to the self-reports of the treatment group members, however, only 48 percent received treatment during this period. Meanwhile 32 percent of control group members self-reported receiving training from other sources over the same interval."* We have to bear in mind that the experimental evaluation results rest on the assumption that treatment in the former group was 100% whereas in the latter it should be 0%. The necessity to invest in implementation evaluation is obviously underlined by this result.

Another aspect which reinforces the case for in-depth implementation evaluation is the increasing acknowledgment of the close interconnection between active labour market policy and the traditional services of the employment service, which include information, counseling, vacancy filling, and also administration of unemployment insurance payments. As it is frequently shown, in many countries reforms of the institutional frameworks of labour market policy are taking place, however these activities are not very strongly accompanied by evaluation research. Some recent studies point to the difficulties, and to the lack of expertise in the evaluation of such a kind of large and complex institutional frameworks.²⁵ Additionally, we can observe that also in major reforms which were recently set in motion, the core idea of implementation research, that a reform is not completed but started by the amendment of a new law, is not taken seriously. As an example we may take the Austrian reform of the employment service, which was complemented by a large scale external assessment of the institution and also of the reform proposals, but during implementation has not been supported by comparable evaluation activities. In-depth implementation evaluation would certainly improve knowledge about these overall structures.

In concluding the section about implementation evaluation we can easily see that this kind of evaluation activity has a close relationship to impact evaluation in at least two respects. First, appropriate structures of monitoring and performance indicators are a necessary precondition for impact studies -- the best methodology to disentangle net effects is not worthwhile, if the data used do not reflect the measures adequately. Second, in-depth implementation analysis often helps to understand the results of impact analysis, and shows us ways how to improve the practices.

²⁴

See as examples the classical implementation studies by Pressman & Wildavsky 1984, or Bardach 1982.

²⁵

As examples see OECD (1989), or the report about the Austrian public employment service (BMAS 1992).

d) Policy evaluation: How to Force the Political System to Make Clear Statements about Goals and Objectives?

In the category of policy evaluation, we are principally confronted with the fundamental question, whether we as evaluators should charge the political system to change its practices significantly, or whether we rather should try to fit into the ongoing political practices in an informed and reasonable manner. The oft-stated request for clear statements of unambiguous goals and objectives from the political community, and the possibility to derive similar clear-cut outcome measures, may be a pious hope rather than as an operative challenge. Nevertheless we can take steps in this direction. In this way we have to acknowledge that in the overall field of labour market policy there are conflicting interests involved, and will always be involved (OECD 1989, 28): one is the efficiency-equity trade-off, another is to which extent services should be allocated to either side of the labour market, etc. Since the proposed new framework for labour market policies was endorsed by labour ministers in 1992, the role actually imposed to active measures of labour market policy is an important question of policy evaluation.

In any case, the point for policy evaluation is that the formulation of goals and objectives should be "internalized" in the evaluation activity, and consequently this dimension should be made open for informed discussion. An obvious and important aspect of policy evaluation, which frequently had not been taken into account, is the analysis of reasonable expectations for an implemented measure: How large could the expected effects be?²⁶

Ex-ante evaluation, which reviews the action plans prior to their implementation, is a useful tool contributing to policy evaluation. Another tool of this kind is the *examination of "evaluability"* which was proposed by Joseph S. Wholey (1979). This kind of exercise has frequently proven to collect and transmit the desired information to program providers, so that it in fact could be used as a substitute for a full range evaluation project.

Concluding Remarks

As increasing resources are spent on activities in labour market policy, the evaluation of these activities is gaining interest in several countries. However, there are large differences concerning the experience in labour market policy evaluation. What can be learned from past experience?

We have stated at the outset, and as a conclusion from the general evaluation literature, that it would be appropriate to complement outcome-oriented impact analysis with a second route of analysis,

²⁶

This question is illustrated by Robert J. LaLonde (1995) in a recent review of training programmes.

which may be termed "process oriented analysis", with emphasis on the implementation of policy measures. The mere observation of what works or what does not work should be complemented by the question why it works or does not.

If we refer to our short account of results, we can easily see that in most categories of measures, there is no simple answer to whether something works or does not work. Rather the outcome seems to depend on the specific nature of measures, which highlights the close linkage of impact and implementation. Moreover, our short comment about outcomes seems to indicate that broad categories of programmes, which cover both a widespread and various populations of clients and a various collection of specific treatments, do not show positive results. This may be because of variability of outcomes, which was not very observable using the traditional methodology oriented to mean effects, or because the measures in themselves lack appropriate targeting to the specific problems of clients. Differentiation of evaluation according to the various areas and categories described will be necessary; however, the more differentiated the practice of evaluation will be, the more it will become difficult to gain an overview from an external standpoint. Therefore it seems to be inevitable in the long run that evaluation concerning all the mentioned categories will be brought to the actors which are themselves providing the policy measures at the "grassroots" level.

In taking steps to develop a comprehensive framework of labour market policy evaluation a combination of external and internal evaluation procedures seems to be the best way to set up such a framework. In this case the chance is high that the drawbacks of both models can be overcome, using the strengths of each model. Internal evaluation procedures are able to bring the evaluation activities in pace with practical needs. In addition, the implementation of proper monitoring procedures as well as of performance indicators can only be performed by internal frameworks. These procedures seem in any case to be a precondition for the performance of external evaluation activities, providing necessary bases for information. External evaluation, especially when it is linked to the academic community, is able to transcend some limitations of the views of insiders, and can also contribute to methodological improvements and ensure proper quality standards.

The different categories of labour market policy evaluation -- policy analysis, impact analysis, and implementation analysis -- should be taken seriously. Priority should be given especially to the development of implementation analysis: setting up sound monitoring systems and performance indicators must be considered as very important steps. Implementation evaluation has a close relationship to impact evaluation because appropriate structures of monitoring and performance indicators are a necessary precondition for impact studies: the best methodology to disentangle net effects is not worthwhile, if the data used do not reflect the measures adequately. Concerning impact evaluation the problems involved in the experimental methodology should be taken seriously into account. Investment in nonexperimental evaluation should be provided because of its

applicability, and because there may also be a way to better combine nonexperimental impact evaluation with the other categories of evaluation, especially with implementation issues.

As a final conclusion we may argue, that there is a movement in the field of labour market policy evaluation going on which follows two joint directions:

- One is a broadening of the focus from programmes to institutions,
- and the second is the movement from the narrow external black-box impact evaluation to the setting up of broad frameworks of self-evaluation, assisted in certain points by external services.

This development may be seen as a process of differentiation and enrichment, which has the figure of blowing into a balloon and seeing how it grows to its full beauty and magnitude -- however, we should blow not too sharp, because in this case ... you know what I mean.

References

- Bardach, E. (1982), *The Implementation Game. What Happens After a Bill Becomes a Law*, Cambridge, Mass: The MIT Press.
- Björklund (1988), What Experiments Are Needed for Manpower Policy, in: *The Journal of Human Resources*, Vol. 23, No.2, 267-277.
- Björklund, A. (1991), Evaluation of labour market policy in Sweden, in: OECD 1991, S.73-88
- BMAS, Federal Ministry of Labour (1992), *Die Effizienz der Arbeitsmarktverwaltung, Organisationsanalyse. Forschungsberichte aus Sozial- und Arbeitsmarktpolitik Nr. 45*
- Breen, R. & B. Halpin (1989), *Subsidizing Jobs: An Evaluation of the Employment Incentive Scheme*, The Economic and Social Research Institute, Dublin.
- Bryant, E.C. & K. Rupp (1987), Evaluating the Impact of CETA on Participant Earnings, in: *Evaluation Review*, Vol.11, 473-92.
- Burtless Orr (1985), Are Classical Experiments Needed for Manpower Policy, in: *The Journal of Human Resources*, Vol. 21, No.4, 606-639.
- Burtless, G. (1995), The Case for Randomized Field Trials in Economic and Policy Research, in: *Journal of Economic Perspectives*, Vol 9 (Spring 1995), 63-84.
- Calmfors, L. (1994), Active Labour Market Policy and Unemployment - A Framework for the Analysis of Crucial Design Features, OECD Economic Studies No. 22, 7-47.
- Chen, H-T. (1990), *Theory-driven evaluations*, Newbury Park: Sage
- de Koning, J. & R.F. de Munnik (1991), Evaluation of Manpower Policy: The Dutch Position, in: OECD 1991, 145-57.
- Delander, L. (1991), *Placement, Counseling and Occupational Rehabilitation in Sweden*, Discussion Paper FS I 91 - 6, Wissenschaftszentrum Berlin.
- Dickinson, K.P., T.R. Johnson & R.W. West, An Analysis of Sensitivity of Quasi-Experimental Net Impact Estimates of CETA Programs, in: *Evaluation Review*, Vol.11, 452-72.
- Employment and Immigration Canada (1989), *Evaluation of the National Employment Service. An Overview Report*.
- Guba, E.G. & Y.S. Lincoln (1987), The countenances of fourth-generation evaluation: Description, judgment, and negotiation, in: *Evaluation Studies Review Annual*, Vol.11, S.70-88
- Hausman, J & D.Wise, eds. (1985), *Social Experimentation*, Chicago: University of Chicago Press.
- Haveman, R. & R.Hollister (1991), Direct Job Creation: Economic Evaluation and Lessons for the United States and Western Europe, in: A. Björklund et al., eds., *Labour Market Policy and Unemployment Insurance*, Oxford: Oxford University Press.
- Heckman, J.J. & J.A. Smith (1995), Assessing the Case for Social Experiments, in: *Journal of Economic Perspectives*, Vol 9 (Spring 1995), 85-110.
- Heckman, J.J., V.J.Hotz & M.Dabos (1987), Do We Need Experimental Data to Evaluate the Impact of Manpower Training on Earnings?, in: *Evaluation Review*, Vol.11, No 4, 395-427.
- Höcker, H. (1994) *Reorganisation der Arbeitsmarktpolitik. Weiterbildung für Arbeitslose in Dänemark*. Discussion Paper FS I 94 - 202, Wissenschaftszentrum Berlin.
- Hofer, H & K. Pichelmann (1995), *Ergebnisorientierte Evaluierung von Instrumenten der experimentellen Arbeitsmarktpolitik in Österreich*. Research Report, Vienna: IHS.
- Jackman, R. (1995), *What Can Active Labour Market Policy Do?* Centre for Economic Performance, Discussion Paper No.226, London: LSE.

- LaLonde, R.J. & R. Maynard (1987), How Precise are Evaluations of Employment and Training Programs? Evidence from a Field Experiment, in: *Evaluation Review*, Vol.11, No 4, 428-51.
- LaLonde, R.J. (1995), The Promise of Public Sector-Sponsored Training Programs, in: *Journal of Economic Perspectives*, Vol 9 (Spring 1995), 149-68.
- Lassnigg, L. et al. (1994), Möglichkeiten der Evaluation und Effizienzbewertung des Arbeitsmarktservice, Research Report, Vienna: IHS
- Lassnigg, L., G. Buzeczki, M. Wiederschwinger, & W. Altzinger (1990), Wirkungen von vermittlungsfördernden Maßnahmen: Betriebliche Einschulung und Verringerung der Lohnkosten, Vienna: Forschungsberichte aus Sozial- und Arbeitsmarktpolitik Nr.42
- Layard, R. (1986), *How to Beat Unemployment*, Oxford: Oxford University Press.
- Layard, R., S.J. Nickell & R. Jackman (1991), *Unemployment*, Oxford: Oxford University Press.
- Linke, L. (1994), Reorganisation der Arbeitsmarktpolitik. Weiterbildung für Arbeitslose in der Bundesrepublik Deutschland. Discussion Paper FS I 94 - 203, Wissenschaftszentrum Berlin.
- Moraal, D. (1994), Reorganisation der Arbeitsmarktpolitik. Weiterbildung für Arbeitslose in den Niederlanden. Discussion Paper FS I 94 - 204, Wissenschaftszentrum Berlin.
- Mosley, H. & Ch. Degen (1994), Reorganization of Labour Market Policy. Further Training for the Unemployed in the United Kingdom. Discussion Paper FS I 94 - 205, Wissenschaftszentrum Berlin.
- Mosteller, F. & D.P. Moynihan, Eds. (1972), *On Equality of Educational Opportunity*. New York: Random House.
- Nathan, R. (1991), Evaluation Strategies With Particular Emphasis on Demonstration Studies, in: OECD 1991, S.21-42
- OECD (1988), *Measures to Assist the Long-Term Unemployed*, Paris.
- OECD (1989), *Manpower measures: Main findings of the evaluation programme*, Paris: OECD (mimeo)
- OECD (1991), *Evaluating labour market and social programmes. The state of a complex art*, Paris: OECD
- OECD (1993a), *The Public Employment Service in Japan, Norway, Spain and the United Kingdom*, Paris.
- OECD (1993b), *The Labour Market in the Netherlands*, Paris.
- OECD (1993c), *Active Labour Market Policies: Assessing Macroeconomic and Microeconomic Effects*. *Employment Outlook*, 39-80.
- OECD (1995a), *The Public Employment Service in Denmark, Finland and Italy*, *Employment Outlook*, 101-38.
- OECD (1995b), *Employment Outlook, Making Active Labour Market Policies More Effective*, Editorial.
- Pichelmann, K. et al. (1991), *Monitoring von AMV-Maßnahmen auf der Basis von Erwerbskarriere-Informationen*, Research Report, Vienna: IHS
- Pressman, J.L. & A. Wildavsky (1973ff), *Implementation. How great expectations in Washington are dashed in Oakland;...*, Berkeley: University of California Press
- Riddell, C. (1991), Evaluation of manpower and training programmes: the North American experience, in: OECD 1991, S.43-72
- Rossi, P.H., H.E. Freeman & S.R. Wright (1979), *Evaluation. A Systematic Approach*. Beverly Hills: Sage.

- Rossi, P.H. & J.D. Wright (1984), Evaluation Research: An Assessment, in: *Ann.Rev.Sociol.* 10, 331-52.
- Scriven, M. (1972), Pros and Cons About Goal-Free Evaluation. *Evaluation Comment*, in: *The Journal of Education Evaluation*, Vol 3 (December), 1-7.
- Scriven, M. (1973), Goal-Free Evaluation, in: E.R. House, ed., *School Evaluation. The Politics and Process*, Berkeley: McCutchan.
- Shadish, W.R., Jr. & R. Epstein (1987), Patterns of program evaluation practice among members of the Evaluation Research Society and Evaluation Network, in: *Evaluation Review*, Vol. 11 No.5 (October), S.555-590
- Stromsdorfer, E.W. (1987), Economic evaluation of the Comprehensive Employment and Training Act: an overview of recent findings and advances in evaluation methods, in: *Evaluation Review*, Vol.11, S.387-394
- Wholey, J.S. (1979), *Evaluation: Promise and performance*, Washington D.C.: The Urban Institute
- Wilensky, H.L. (1985), Nothing Fails Like Success: the Evaluation-Research Industry and Labour Market Policy, in: *Industrial Relations*, 24 (Winter), 1-19.
- Zilian, H.-G. & B. Malle (1993), *Spreu und Weizen. Das Verhalten der Arbeitskräftenachfrage*, Research Report, Graz: Büro für Sozialforschung

Institut für Höhere Studien
Institute for Advanced Studies

Stumpergasse 56
A-1060 Vienna
Austria

Phone: +34-1-599 91-216

Fax: +34-1-597 06 35