

EMPLOYMENT AND TRAINING  
PAPERS

49

# A framework for active labour market policy evaluation

Gaëlle Pierre  
Department of Economics  
University of Warwick

Employment and Training Department  
International Labour Office Geneva

ISBN 92-2-111790-1  
ISSN 1020-5322  
First published 1999

## Foreword

In this paper the author undertakes a comprehensive survey of the theoretical and empirical literature on the evaluation of active labour market policies (ALMPs). The contribution of the paper is to provide more accessible and practical guidelines for policy evaluation targeted at a wider audience, including policy makers and practitioners.

ALMPs rather than passive labour market policies have increasingly been promoted in the OECD countries and transition economies as a principal means to deal with unemployment. In the United States and Britain, ALMPs known as welfare to work or workfare programmes have become the dominant strategy to bring the jobless back to work. Among the EU countries, activation of labour market policies aimed at the creation of a high-skilled and knowledge-based economy have become the key concept behind the European Employment Strategy. The evaluation of ALMPs should not be considered as a one-off process. It should be used to improve the next phase of policy design. Evaluation is part of the dynamic process of improving policy design, implementation, monitoring as well as evaluation techniques.

The paper defines the main issues to be examined regarding the evaluation of ALMPs, provides non-technical explanations for various evaluation techniques, and assesses their strengths and weaknesses. Having introduced some examples of applying different techniques, the author assesses their practical use. By examining cases in the European Union and the United States, the author stresses the need for a systematic approach of policy evaluation as well as for the effective use of the evaluation results for the future policy making and implementation of ALMPs.

Finally, the author also identifies the target group approach as the most effective for evaluations, and argues that a combination of evaluation techniques should be applied. Monitoring, transparency, a careful use of data also improve the quality of evaluations. Most empirical literature focuses only on the employment and income effects of the ALMPs, but the author points out that the social and equality aspects are equally important for the evaluation of labour market policies.

Gek-Boo Ng  
Chief  
Employment and Labour Market Policies Branch

## **Acknowledgements**

I wish to thank Peter Auer, Niall O’Higgins, June Ishikawa, Sandrine Cazes, Jennifer Smith, anonymous referees and my supervisor Gek-Boo Ng for providing helpful advice and comments on earlier drafts.

## Table of Contents

Foreword .....	i
Acknowledgements .....	ii
1. Introduction .....	1
2. What does policy evaluation involve? .....	3
2.1. The effects of active labour market policies on the functioning of the labour market .....	3
2.1.1. Direct effects on employment and unemployment .....	3
2.1.2. ‘Secondary’ effects .....	4
2.1.3. Interaction of ALMPs with other policies and institutions .....	5
2.1.3.1. The identification of ALMPs’ impacts .....	5
2.1.3.2. Institutions and ALMPs’ formation and effectiveness .....	5
2.2. Evaluation versus monitoring .....	6
2.3. The implementation of the evaluation .....	7
2.3.1. The evaluators .....	7
2.3.2. The timing of evaluation .....	7
2.3.3. Data requirements .....	8
2.3.3.1. Survey data .....	8
2.3.3.2. Administrative data .....	8
2.3.4. Generalization and use of evaluation results .....	9
3. The main techniques .....	10
3.1. Quantitative analysis at the microeconomic level .....	10
3.1.1. Quasi-experimental analysis .....	10
3.1.1.1. Definition .....	10
3.1.1.2. Strengths and weaknesses .....	11
3.1.2. Experimental analysis .....	14
3.1.2.1. Definition .....	14
3.1.2.2. Strengths and weaknesses .....	15
3.2. Aggregate impact analysis .....	16
3.2.1. Definition .....	16
3.2.2. Strengths and weaknesses .....	17
3.3. Cost benefit analysis .....	18
3.3.1. Definition .....	18
3.3.2. Strengths and weaknesses .....	21
3.4. Qualitative analysis .....	22
3.4.1. Definition .....	22
3.4.2. Strengths and weaknesses .....	22
3.5. Lessons for determining a framework for evaluation .....	23
4. A framework for evaluation .....	24
4.1. A target-oriented approach .....	24
4.2. Policy formation stage .....	25
4.3. Policy implementation stage .....	27
4.4. Impact stages .....	27
5. Conclusion .....	30
Appendix: The Layard-Nickell model .....	32

- Box 3-1. Assessment of the net effects of Employment Training and Employment Action using matched comparison method
- Box 3-2. Impact of youth employment schemes on subsequent employment and unemployment durations
- Box 3-3. The Job Training Partnership Act (JTPA)
- Box 3-4. The role of labour market policies and industrial settings on unemployment: A cross-country study
- Box 3-5. Macroeconomic effects of ALMPs
- Box 3-6. Cost benefit analysis of JTPA study
- Box 3-7. Cost benefit evaluation of the job corps
- Box 3-8. A qualitative study of the “Investors in People” programme
- Box 4-1. The evaluation of ALMPs in practice: The French example
- Box 4-2. Examples of combinations of techniques
- Box 4-3. An example: Impact evaluation of policies targeted at youths

# 1. Introduction

Faced with a large increase in unemployment over the past twenty years (see table 1 for current unemployment rates in the European Union (EU) and the United States) and the realisation that aggregate demand policies were not sufficient to solve the problem, industrialized countries have developed policies aimed at reducing labour market imperfections and at preventing the degradation of the situation of some disadvantaged groups of people (the youth, disabled, unskilled, long-term unemployed, etc.). These active labour market policies (ALMPs) can be divided in three categories: labour market training, direct job creation, and job brokerage (i.e. improving the match between job seekers and vacancies) (Calmfors, 1994).

Although ALMPs are not sufficient to solve the unemployment problem and although they are sometimes considered to be ineffective (in view of the mixed evidence gathered on their effects), this paper takes the view that they are useful. The recent persistent unemployment problem has indeed created specific social issues such as exclusion (from the labour market and from society). Several authors (Gregg and Wadsworth, 1996) have observed a parallel increase in the number of households in which no adults work and households in which all adults work. Moreover, the current difficulties faced by many groups, such as single mothers, may be translated into persistent social exclusion for the future generations. Insofar as ALMPs are an attempt to alleviate these inequalities, through a better integration in the labour market and a fairer income distribution, their use is justified.

Table 2 shows that in most EU countries a significant percentage of GDP is devoted to ALMPs. Given the recent increase in their use (average ALMP spending in the EU went from 0.72 per cent of GDP in 1985, to 0.85 per cent in 1989 and 1.10 per cent in 1994/95 (Auer, 1997)), and the question mark attached to their effectiveness, their evaluation is crucial. An evaluation will aim at informing governmental decision for designing future ALMPs, but also for improving the quality of current policies. It will also provide information for a public debate to take place, and more generally for individuals and firms who participate in programmes as well as those who do not.

ALMP evaluation has not been undertaken systematically. Where it has occurred, the aims and uses of evaluation have depended on the political and institutional contexts of the country considered. Surprisingly, the United States has been carrying out evaluation extensively while it spends relatively little money on ALMPs, while the EU countries have few evaluation studies and are spending comparatively more money. A striking example is Sweden which spends a comparatively large share of its GDP on ALMPs (see table 2) but where evaluations are rare and often concern a small sample of the population (Gautié et al., 1994).

The aim of this paper is to address the issues involved in policy evaluation (section 2), to review current practices in OECD countries (section 3), and to determine a framework for future evaluations (section 4). In order to define such a framework, section 2 starts with an outline of the effects of ALMPs on the labour market, then clarifies the concept of programme evaluation (in particular making the distinction between evaluation and monitoring) and finally outlines the main issues of the implementation of evaluation. Section 3 reviews the literature concerning the main techniques used in evaluation and section 4 offers a practical guide to evaluation.

**Table 1. Unemployment rates in 1997, EU and USA.**

Austria	Belgium	Denmark	Finland	France	Germany	Greece	Ireland	Italy	Lux. <sup>a</sup>	Netherlands	Portugal	Spain	Sweden	UK	USA
5.2	9.0	5.4	14.5	12.4	9.8	9.8	10.5	12.5	3.3	5.6	6.9	20.9	8.0	7.1	5.0

a. 1996

Source: OECD employment Outlook 1998

**Table 2. Public expenditures on ALMPs as a percentage of GDP in 1997, EU and USA.**

	Austria	Belgium <sup>a</sup>	Denmark	Finland	France <sup>a</sup>	Ger.	Greece <sup>a</sup>	Ireland <sup>a</sup>	Italy <sup>a</sup>	Lux. <sup>a</sup>	Neth.	Port. <sup>a</sup>	Spain	Sweden	UK	USA
Public employment	0.13	0.22	0.13	0.14	0.16	0.21	0.11	0.24	0.04	0.03	0.35	0.11	0.08	0.26	0.18	0.06
Training	0.09	0.28	0.97	0.55	0.36	0.36	0.07	0.21	0.01	0.01	0.13	0.31	0.14	0.43	0.09	0.04
Youth programme	0.06	0.03	0.10	0.22	0.24	0.07	0.03	0.24	0.42	0.14	0.10	0.34	0.07	0.02	0.13	0.03
Subsidized employ.	0.21	0.84	0.31	0.53	0.48	0.34	0.05	0.88	0.61	0.06	0.42	0.12	0.20	0.70	-	0.01
Disabled measures	0.06	0.14	0.28	0.12	0.08	0.28	0.01	0.08	-	0.04	0.54	0.07	0.02	0.67	0.02	0.03
Total ALMPs	0.44	1.50	1.79	1.57	1.32	1.25	0.27	1.66	1.08	0.27	1.53	0.95	0.50	2.09	0.42	0.17
Total <sup>b</sup>	1.73	4.27	5.80	4.79	3.13	3.79	0.71	4.07	1.96	0.95	4.86	1.97	2.37	4.25	1.47	0.43

a. 1996

b. Total = passive + active labour market policies.

Source: OECD Employment Outlook 1998

## 2. What does policy evaluation involve?

### 2.1. The effects of active labour market policies on the functioning of the labour market

In order to understand what needs to be evaluated, it is necessary to outline the effects that active labour market policies have on the functioning of the labour market. ALMPs are expected to correct labour market imperfections. They act in various ways that are reviewed in this section.

It is useful to note here that the terms “policies” and “programmes” are used interchangeably in this paper. They however represent different concepts. The first is a generic term representing the broad direction that the government wishes to take with respect to specific economic issues. For example, the government may choose to use active labour market policies rather than passive labour market policies, or it may decide to increase the flexibility of the labour market, etc. Programmes are specific measures taken in order to carry out these policies, such as wage subsidies or classroom training.

#### 2.1.1. *Direct effects on employment and unemployment*

Labour market policies are expected to facilitate the matching between workers and employers. For instance, retraining programmes are expected to give workers the skills required by firms, whilst employment services can provide better information on vacancies or help to improve the search effectiveness of the unemployed. Other labour market programmes can be substitutes for regular work and maintain or restore the employability of the participants. However, the evaluation of these effects is not always straightforward. For example, one has to take into account the locking-in effects (participants in the programme have less time to look for a regular job) which have to be subtracted from the positive effects once the programme has been completed. These locking-in effects may be expected to be quite small when the programme targets people who are already hard to place and have a record of non-placement. The perspective of participating in a programme may also decrease the incentive to find a job.

In a macroeconomic evaluation, interested in the employment effect of a programme, it may also be useful to note that participation in a programme may automatically reduce the number of claimants, but not the actual number of unemployed. This occurs in cases where programme participants are not counted as unemployed but are still looking for a job; ‘measured’ unemployment is reduced while ‘actual’ unemployment is unchanged. It is therefore necessary to keep such possible bias in mind when interpreting results from macroeconomic studies.

Unemployment tends to reduce labour force participation; discouraged or older workers may leave the unemployed pool or some people may not enter at all (women for example). In giving opportunities to unemployed persons, ALMPs can maintain participation in the labour market. These new opportunities also give incentives to people who were out of the labour force to enter the labour market. By maintaining or inducing participation of the labour market, ALMPs cause an increase in the size of the labour force. A larger labour force has two effects. On the one hand, as the labour force expands and labour demand is unaffected (firms still employ the same number of workers at the going wage rate), a lower proportion of the labour force is regularly employed, hence a larger proportion is unemployed or participating in programmes. On the other hand, the increase in labour supply relative to demand creates more competition in the labour market so that the wage rate falls. This fall



leads to new jobs being created and to a larger proportion of the population as a whole being regularly employed.

In addition to this ‘crowding in’ effect, programmes like job subsidy schemes, which are often targeted at some group among the unemployed, may have particular effects on the rest of the population. Three complex and contradictory effects have been identified (Calmfors, 1994; Schmid, 1996). The first one is the dead-weight loss; this loss exists because a certain proportion of the hirings that take place under the programme would have taken place anyway. The second is the substitution effect, which is the fact that job created under the subsidy may replace jobs for other categories of workers. These problems are present when the additionality principle (any created job should be a new one) is not imposed. The third effect is called displacement effect: firms which benefit from the wage subsidies may gain competitive advantage in the product market and increase their share of the market at the expense of other firms which may have to dismiss workers.

### 2.1.2. ‘Secondary’ effects

One of the aims of ALMPs is to improve the relative social situation of the unemployed, although the interest given to this aspect varies across countries and time. For example, compensation given in programmes may be higher than unemployment benefits. Participation in programmes may also lower the risk of future unemployability or unfavourable career development due to protracted periods of unemployment (Calmfors and Nymoen, 1990; Calmfors and Lang, 1993). However, making unemployment a less difficult experience may have spurious effects, in particular a positive pressure on wages. The latter is due to the fact that the threat of unemployment has been reduced. Workers or unions will have more power in wage bargaining, and their wage claims will be higher (Layard et al., 1991; Shapiro and Stiglitz, 1984). This increase in wages may also lead to the destruction of jobs in the rest of the economy; this is another form of the displacement effect.

The productivity of workers is supposed to deteriorate when they are idle. ALMPs, such as training or even job creation programmes, can therefore raise or maintain the productivity of unemployed persons. The employment and wage responses to this type of programme is however not clear. The increase in productivity which is induced may have two opposite effects on employment: firms may have an incentive to expand output by employing additional workers (who cost less per unit of output), but they may also reduce labour demand as they need less workers to produce the same output. Moreover, if one believes that part of the increase in unemployment has been due to the lack of downward adjustment of wages to the slowdown in productivity, unemployment can also be reduced.<sup>1</sup> However, this effect is likely to be rather small. These programmes can also be seen as a path for workers from low-productivity sectors (with excess supply) to high productivity sectors (with excess demand). One can show that when people are “transferred” from the low productivity to the high productivity sector, the employment rate must increase (see appendix for a detailed explanation). This argument is less powerful in a situation of general excess supply of labour.

ALMPs enable the authorities to check whether people are genuinely looking for work. This will reduce the claimant count of unemployment, since only genuine unemployed people continue signing on; however, this has by definition no effect on involuntary unemployment.

<sup>1</sup> Some authors (Sachs, 1987) believe that part of the increase in unemployment in the past two decades was due to the fact that productivity did not increase by as much as wages, so that firms were faced with greater unit labour costs.

ALMPs may also have tax repercussions.<sup>2</sup> The first possible effect is the following. By increasing employment, they increase the tax base so that the total costs of unemployment benefits and of the programmes are lower. This should lead to a decrease in the tax rate. In turn, firms will employ more workers at each level of after tax wage (the wage received by the workers), because they pay a lower pre-tax wage. If this effect is not fully compensated by an increase in the real wage claimed by the workers at each level of employment, total employment will increase. Other effects may work in different directions: ALMPs may be more costly than unemployment benefits (or more people may claim benefits as a result of the implementation of ALMPs), in which case, the tax rate will tend to be raised. This will have the opposite effect to the one just seen and employment will decrease. However, a larger labour force participation may also reduce the costs for early retirement pensions (and decrease tax rates). If it reduces the number of claimants, the work test effect (explained above) may also reduce cost.

### ***2.1.3. Interaction of ALMPs with other policies and institutions***

#### *2.1.3.1. The identification of ALMPs' impacts*

The net effects of ALMPs may be difficult to identify because of their interaction with other policies. These other policies include passive labour market policies, demand management policies or structural policies of the labour market (minimum wages, wage-setting practices, employment protection legislation, etc.). All these may affect labour market outcomes of unemployed people. This is particularly relevant if they have been altered at the same time as the ALMPs were implemented. This is particularly crucial in aggregate studies, which try to explain the net effect of programmes on the economy as a whole.

#### *2.1.3.2. Institutions and ALMPs' formation and effectiveness*

Institutions in a broad sense<sup>3</sup> are important determinants of how policies are formed and of their effectiveness. Schmid and Reissert (1988 and 1991) and Freeman (1998), among others, outline that institutions are highly country specific and thus can be considered as one of the reasons for cross-country differences. In the labour market, the effects of institutions, such as the unions and work councils or government regulations, on unemployment and wages have been frequently studied in the economic literature. They are relevant to policy evaluation, in particular if one wants to do cross-country comparisons or to study a country over time. The application of the same policy may have different effects depending on the country in which it is implemented (Freeman, 1998).

We do not review all possible interactions between Institutions and labour market policy formation, but we report an interesting example developed in Schmid and Reissert (1988 and 1991). They explain how the choice of policy response to soaring unemployment depends, among other things, on the way financing institutions work. They find evidence in a cross-country study that financing arrangements affect various aspects of active labour market

<sup>2</sup> The extent of these repercussions will depend on the policy-makers' will and on the institutional structure of the country. For example, an increase in income tax revenue may not always translate into a decrease in tax, depending on the priorities of the policy makers.

<sup>3</sup> Defined as norms, habits, traditions, organisations, procedural and financing regulations, and the distribution of responsibilities and jurisdiction (Schmid and Reissert, 1991).

policies. Their main findings show that the choice between contribution based and tax based financing affects the volume of expenditures for ALMPs and the dynamics of expenditure trends (such as the ability of rapid counter-cyclical response, or the tendency to continuous spending growth). They also show that the share of ALMPs in labour market policies depends partly on the co-ordination between institutions that finance the measures and institutions that financially benefit from them. Inefficiencies can occur in cases where financing institutions have no incentives to choose the least costly strategies, because they do not get themselves the benefits.

It may be difficult to formally and quantitatively evaluate the effects of these interactions. They can however be taken into account thanks to qualitative analysis.

## **2.2. Evaluation versus monitoring**

This section distinguishes between monitoring and evaluating the implementation of the programmes and policies. In practice, these concepts are often confused or considered to be the same. It is however useful to make a clear distinction between the two, as they do not have the same use (Auer and Kruppe, 1996).

Monitoring judges the gross outcomes of the policy against set targets. These include the following: the number of participants from the target group, the expected cost of the programme, the completion rate, the employment status and the qualification reached after participation. Monitoring is considered as a useful exercise to assess the success or failure of a programme, but it does not provide explanations for it. It can be a form of control over the agents implementing the programme, even though in practice few EU countries give incentives to their staff and none apply sanctions. Its main advantage is that it provides rapid information about the programme (as opposed to evaluation, which is often a longer-term process). It provides a useful feedback for improving the implementation of the monitored programme and it directly helps the administrators in their work. Overall, in the countries considered here, where the administration of programmes is usually decentralized, monitoring is considered to be necessary for an efficient delivery of services.

By assessing the policy implementation and outcomes, evaluation aims at determining why a programme is successful or not. Evaluation can take place at all stages of the programme: formation, implementation and outcomes. What is called process evaluation covers the formation and implementation stages (Schmid, 1996). Process evaluation will use the information provided by the monitoring exercise, but its aim will be to explain why targets are not reached. It might for example use a qualitative analysis (see section 3.4) to answer this question. At the outcome stage, one talks of impact evaluation. The effects of the programme are measured and compared to what would have happened without the programme. It is expected to provide feedback for future policies.

Monitoring is a simple and practical mechanism in policy making. It could be considered as a minimum requirement to check how large sums of public money are spent. The information it provides can be used by policy evaluators so that monitoring may be considered as a complement to evaluation. However, monitoring is not sufficient in itself to have a complete view on the effectiveness of the ALMPs because it does not take their secondary effects into account.

## 2.3. The implementation of the evaluation

### 2.3.1. *The evaluators*

It is important to determine who is best qualified to carry out the evaluation. This section makes here the distinction between internal and external evaluation. Although some of the public labour market administrations run effective evaluation branches, an external evaluation is generally considered to be more accountable, as it is supposed to be more objective (Fay, 1996). For example, if the institution in charge of the evaluation is also in charge of the implementation of the programme, one might think that it will have an incentive to find results that correspond to its own agenda. As we will see later, evaluations require certain choices; these include the choice of which effects of programmes' participation to study, how to go about their estimation, etc. Compared to internal branches, which are bound to follow the 'culture' of the administration they depend on, external evaluators will bring an independent and new point of view on these issues. Moreover, internal evaluations might tend to focus on the interests of the political party in power, while external evaluations may be more able to follow an independent benchmark.

In practice however, not all countries systematically use independent consultants to evaluate active labour market programmes. Many evaluations are done by agencies that depend directly on ministries or sometimes by the employment agencies themselves. This can be seen for example for France in the review of evaluation exercises over the past 30 years published by an agency from the labour ministry DARES (1996). It is therefore useful to outline here how such "in-house" evaluations can be improved. First, in order to make these evaluations more accountable, transparency is very important. To reach this, information concerning the vital statistics useful for evaluation can be published and evaluation reports can explain how they are measured and used. It is also necessary that sufficient statistics be produced so that independent researchers can form their own ideas concerning the efficiency of programmes. Moreover, the governmental agencies or departments that are responsible for these evaluations have to be reachable by outside researchers, and provide direct answers to the public. External bodies can also undertake some parts of the evaluation process. For example, the independent National Institute for Economic and Social Research is responsible for estimating the extent of displacement and substitution in the New Deal programme<sup>4</sup> in the United Kingdom. Second, clear objectives have to be attached to the programme. We will see that this is a requirement for efficient evaluation, but it is also important if one wants to improve "in-house" evaluations. With clear aims, the public as well as researchers may be better able to judge whether the programme has been successful.

In allowing the government evaluation to be checked and challenged, accountability would be increased and as a consequence, through a feedback process, evaluation by internal bodies would be improved.

### 2.3.2. *The timing of evaluation*

As previously mentioned, evaluation should not be restricted to explaining the programme's impacts. Process evaluation takes care of the formation and implementation stages and impact evaluation takes care of the effects of the programme.

The timing of the impact evaluation is also crucial in order to have a complete view on

<sup>4</sup> The New Deal aims at improving the prospects of long-term unemployed and young people.

the policy. One needs in particular to distinguish between the short-term and the long-term effects of the programmes. The evaluation of ALMPs is generally carried out over several months or even years after the programme has been implemented. The evaluation period is an important dimension as the impacts of a programme are generally spread over time. The interpretation of any evaluation results has therefore to take into account this time dimension. Given this long-term dimension, it makes sense to use longitudinal data to carry out the estimation, i.e. variables or individuals observed over a certain period of time (if possible covering the time before, during and after the programme which is evaluated). Some authors advocate this longitudinal dimension as the way forward and the solution to evaluation problems, in particular to the various biases and hidden effects involved (see for example Schömann, 1996). However, others such as Heckman (1996) argue that longitudinal studies merely allow different and not necessarily simpler or more plausible assumptions to be made. In spite of these statistical considerations, it has appeared in recent literature that longitudinal data are of great interest to those who study individual labour market outcomes. Overall, they consequently appear to be very relevant and useful for policy evaluation.

In addition to identifying the short-term and long-term effects, it is important to determine the stage at which the implementation of the programme is. If a programme is evaluated too early, whilst not all services are in place, the impact evaluation may give a wrong result (Ryan and Grubb, 1998). No time limit can precisely be set on policy implementation, but the idea is to take into account the fact that some policies may take more time than others to become fully effective.

### **2.3.3. Data requirements**

As far as the data are concerned, evaluators face usual problems linked to the particular kind of data they use.

#### *2.3.3.1. Survey data*

Survey data (based on interviews with individuals) suffer from various problems. When respondents are asked to relate events which have happened in the past, they may not remember the events precisely and for example may underestimate the duration of unpleasant experiences such as unemployment, i.e. the data may suffer from a recall bias. One way to deal with this problem is to also interview a parent or a friend in order to crosscheck the answers. Moreover, it is always difficult to ask people about what they think or what they feel. People are not always consistent in their judgements, and the answers to particular questions may vary with their mood or situation. These data must be used with great care, and it is generally necessary to ask people the same thing at different points in time. These surveys also suffer from non-response, which may possibly create a bias if the people who do not respond share the same characteristics. In the case in which people are followed over a long period of time, surveys may also be subject to attrition, i.e. the decrease in the number of respondents. This can again create a bias because people who drop out of surveys may share specific characteristics. In addition to these issues, surveys require a large financial investment to be set up (the design of the questionnaire is crucial and has to be done with care) and carried out (many interviews have to be arranged, requiring a large staff to do them).

#### *2.3.3.2. Administrative data*

Administrative records provide a useful source of microeconomic data. They are generally reliable and free of sampling error. For example, precise information can be obtained

on the labour market experience of workers, who, as we have just mentioned, might not be able to remember clearly all the events in their life. However, these data are generally not meant for research and only provide limited information. Studies which use such data are generally restricted in the number of questions they can investigate. Moreover, for security or legal reasons, their use and the combination of data coming from different administrative sources are often restricted. In the case of evaluation studies, researchers may be unable to evaluate fully the programme (for example, they may have information only on wages and not on employment outcome), or may find it difficult to take certain effects into account (such as the displacement or substitution effects) because of insufficient information. In cases where administrative data have been matched with data from other sources, they have proved to be very useful.

Macroeconomic indicators may be subject to measurement errors, or to changing definitions. The measure of the unemployment rate is subject to frequent redefinition in an attempt to improve it, but sometimes also to hide certain unemployed people (placing them in 'other categories'). The solution to this type of problem can be to use standardized and commonly accepted data, such as those published by the OECD or the ILO, to use several different series which are supposed to measure different aspects of the same thing,<sup>5</sup> or even to use different sources for the same series. However, available macroeconomic data may simply be insufficient, although this should not be the case for most EU countries and for the United States. In general, the interpretation of results of macroeconomic studies should take into account any possible bias due to the way series are measured. A general limitation of macroeconomic data is that they cannot capture change caused by particular ALMPs or the effects on individuals. They should therefore not be used in isolation, unless one is interested only in the overall economic outcome of ALMPs.

#### ***2.3.4. Generalization and use of evaluation results***

Evaluation results are supposed to be useful for future policy formation. However, individual evaluations are often restricted to special cases or groups of people. This is the issue of external validity: the results of most evaluations, in particular of microeconomic studies cannot be generalized.

The feedback process is also limited by several problems. First, there must be a political willingness to use evaluation results in future policy formation. This political will may not always be present, especially when there is a change of government and the new policy interests are different. However, this is a two-way relationship; evaluation reports, prepared by experts in the area, may also be totally unreadable by the people who are supposed to use them. If evaluators want to be useful, they have to keep in mind that the policy makers will have to make decisions in the end: they will have to simplify their reports, or at least make clear recommendations. Second, the success of the feedback will also depend on the results obtained: evaluations will fail to lead to improvements in future programmes if they do not answer the interests and concerns of decision-makers (Bonar Blalock, 1990).

The results provided by policy evaluation may also be used to fit political ideologies. Ryan and Grubb (1998) explain how evaluation results have been blatantly ignored by governments (which had ordered them in the first place). Decisions are often made on political

<sup>5</sup> For example, the employment rate, participation rate, vacancy rate and unemployment rate can give a picture of the situation of the labour market.

grounds, with or without evaluation taking place. According to Ryan and Grubb (1998), it seems however that consistent and repeated evaluation results eventually lead to corresponding changes in policies. More generally, the issue of the use of evaluation in policy, like that of the use of social research as a whole, is not a simple one. Pointing out the importance for the researchers to improve their contribution to policy is as relevant as pointing out the need for politicians and administration to use research in policy making (Weiss, 1979).

### 3. The main techniques

#### 3.1. Quantitative analysis at the microeconomic level

The basic idea of the following two techniques is to compare the average (mean) outcomes of the participants in the programme with the outcome of the non-participants (these form two groups: the “treatment group” who participates in the programme and the “control group”, those who do not participate in the programme, but share some characteristics with the participants). The outcomes considered are generally future earnings and employment. The two techniques differ in the way these groups are constructed.

##### 3.1.1. *Quasi-experimental analysis*

###### 3.1.1.1. *Definition*

Quasi-experimental analysis is generally carried out after the programme has been implemented for some time. It uses data collected on individuals, either through administrative records or through interviews (or both).

In this type of analysis, the treatment group includes the participants in the programme, while the control group is composed of persons who have not participated in the programme and who have the same characteristics as the participants. Different methods have been used to select the control group: these can be individuals who dropped out of or were turned away from the programme, individuals belonging to the target group of the programme but who did not apply, individuals from areas outside the programme’s coverage (in the case of local pilot approaches), the participants’ experience before the programme (before/after comparison) and non-participants drawn from other micro-datasets. Some of these studies also compare the outcome of the treatment group with a group of persons with different characteristics (another age group, the opposite sex, the participants in other programmes, and other countries). The latter studies are sometimes called “weakly experimental evaluation” (Ryan and Grubb, 1998).

The evaluation generally takes the form of a statistical analysis, but it can also consist of simple comparisons of outcomes. In the case of statistical analysis, the outcome is the variable that has to be explained (for example: earnings some time after the completion of the programme). The outcome is explained by a set of variables representing factors external to the individuals but which can affect their wage (for example industry, region, etc.), individual characteristics (age, education, work experience, etc.) and a variable indicating whether they participated in the programme or not. The measured influence of this last variable represents the mean effect of participating in the programme on the outcome. The participation variable is also influenced by external variables, which may be the same as the ones determining the outcome and which may or may not be observable by the evaluator.

### 3.1.1.2. *Strengths and weaknesses*

All of the studies mentioned above encounter specific problems, in particular concerning the quality of the control groups. For example, it is thought that, due to their very action, the people who dropped out of a programme are significantly different from the participants.

The main issue in quasi-experimental analysis (an issue which is also addressed in experimental evaluation) is selection bias, which may affect the accuracy of the estimates. Selection bias means that a better outcome for the participants compared to the non-participants may be due to differences in the characteristics of the persons in the two groups and not to participation in the programme. For example a large difference in the earnings of participants compared with those of non-participants may be due to the fact that more motivated and able people participated in the programme. If the selection bias is not corrected in the statistical analysis, the evaluators will overestimate the effect of the programme on participants. This selection bias can be due to observable or to unobservable variables.<sup>6</sup>

If the variables causing selection into the participating group are observable, the way to correct the bias is to include these variables in the estimation. A popular way to do that is to use an “external” comparison group (drawn from a different dataset) constructed by matching the individuals of this group to the individuals of the treatment group according to characteristics that are supposed to influence participation (age, gender, education, unemployment experience, etc.). Several recent examples are available in the literature; one of these is detailed in Box 3-1.

<sup>6</sup> One distinguishes between observable variables, which can be seen and measured by the evaluators, and unobservable variables, which cannot.



**Box 3-1. Assessment of the net effects of Employment Training and Employment Action using matched comparison method (Payne et al., 1996, United Kingdom)**

Employment Training (ET) was the major UK government programme of training and work experience for adult long-term unemployed up to 1993. Employment Action (EA) was a smaller programme for adult long-term unemployed, with an emphasis on work placements. The programmes were replaced in 1993 by Training for Work, which combined features of both. (A new policy called “welfare to work” is now being implemented.)

The evaluators used a personal interview survey of participants in the programmes and a matched comparison group of non-participants. The interviews of the participants took place soon after the end of the programme and again one year later, while the non-participants were interviewed only in the second instance. There was a final follow up, a year later. The matching of the comparison group was done using the following characteristics: geographical location, gender, age, and period of unemployment (start and duration). They used both statistical analysis and outcome comparisons.

This type of analysis enabled only the comparison of the participants with non-participants who have the same characteristics as them. In particular, the outcome of the participants was not compared with the outcome of all other unemployed people, only with the outcome of people with the same unemployment experience as them.

They find evidence that ET had a positive impact on the chance of getting a job, while the evidence for the effect of EA was not clear. There was no evidence that wages were higher. The evaluators were also able to determine which particular services had most effects within the programmes. They also analysed the data in order to determine whether the participants were different in some way from the non-participants.

The evaluation was useful in answering the particular questions asked by the evaluators. They were not able to have a complete picture on the effect of the programme on the labour market as a whole (as they did not identify the displacement and substitution effects). They however made a qualitative judgement on the possible extent of these effects. The economic activity at the time of the programme was also thought to have an effect on the success on the programme, but the direction of this effect could not be evaluated. The attrition of the data (i.e. the fact that people had dropped out of the study, especially in the last follow up) created some biases, which they took into account. Overall, the evaluation was limited to the particular type of individuals studied and could not be generalized, but was quite successful in answering the questions it was asked.

Where selection bias is due to unobservable factors, the bias is less easily corrected. Evaluators have to make assumptions about the relationship between participation in a programme and some unobservable factors. The main weakness of such technique is that the results are strongly dependent on the assumption made about this relationship, so that very different results may be obtained without ever being able either to choose or even to have a meaningful comparison between them. Warburton (1996) gives an example of a programme implemented in the United States, which was found to have opposite effects by different researchers using the same data. Warburton manages to identify the cause of their

disagreement, which he attributes to the type of data used.<sup>7</sup> Among the assumptions and methods often used, the following will be explained: fixed-effects model and instrumental variables.

The first method assumes that the unobservable factors which influence programme participation differ across individuals but do not vary over time (for example each individual has some personal and constant view on labour market programmes and will always (never) participate if possible). The problem is that this assumption is often too strong (Heckman and Smith, 1996).

The second method involves finding an “instrumental variable”. This variable must be strongly correlated with the participation decision of the people, but it must be independent of any unobservable variable that also influences the outcome. As no unobservable factors influence it, its measured effect on the outcome is considered to be a good estimate of the effect of the participation to the programme. This type of model is sensitive to the assumption about the decision process of entering into a programme or not. To be successful, it requires that, in cases where the response to the treatment varies, individuals’ gains from the programme that cannot be predicted from variables available to observing social scientists do not influence the decision of the person being studied to participate in the programme. In practice, finding a good instrumental variable may prove difficult (Heckman and Smith, 1996). An example of a study with quasi-experimental data is presented in Box 3-2.

Another type of bias may be present in such analysis as well as in experimental analysis: substitution bias. It comes from the fact that alternative programmes, which can be close substitutes to the programme studied, may be available to the control group. If members of the control group do participate in other programmes, their labour market outcome is affected and, in effect, the alternative to ‘participation in the programme’ is not clearly defined.

<sup>7</sup> The evaluated programme provides training. The outcome considered is the subsequent income of the participants. The discrepancies in the results of different studies are thought to come from the fact that the authors of the studies use annual earnings data and ignore unemployment insurance and welfare dependence.

**Box 3-2. Impact of youth employment schemes on subsequent employment and unemployment durations (Bonnal et al., 1997, France)**

The data used were collected by INSEE. They consist of a survey in which unemployed people, randomly drawn from the files of the public employment services, were interviewed four times over the period 1986-1988. The data provide information on their skill level, education, reason of entry into the sampled unemployment spell and individual characteristics (such as qualification for unemployment insurance, previous participation in a programme, duration of the sampled unemployment spell). They use the data concerning men who were younger than 26 in November 1986 and who obtained either a technical school certificate or no diploma at all.

The authors model the transitions of individuals among six different labour market states (unemployment, regular employment with a permanent contract, temporary employment with a fixed-term contract, employment in a public employment programme, out of the labour force and attrition<sup>8</sup>).

They then analyse whether the previous occurrence of an employment programme<sup>9</sup> has any effect on future transition intensities. They obtain three types of results: they are able to determine the effect of each type of programme on future working life, they identify what factors determine participation in programmes and finally they outline the effect of the duration of the period of entitlement to unemployment insurance on the expected duration of unemployment spells.

Their analysis provides valuable information on the effectiveness of the programme, but is again limited to particular questions. It is also limited by the information available in the data. In particular, their results cannot be generalized, and do not give an idea of the impact on the economy (or, for example on young workers as a group); this is again due to the fact that they cannot estimate the displacement effect.

### **3.1.2. Experimental analysis**

#### *3.1.2.1. Definition*

In experimental analysis, the evaluation is based on an experiment. A randomly selected sample of people is formed, and the difference between the mean outcomes of the participants and the non-participants is supposed to be the net effect of the programme (cleared of biases). The two groups of people are constructed in the following way: people filling the selection criteria of the programme are first selected (pre-selection), then some of them are selected randomly to participate in the programme while participation is refused to the others.

Experimental analysis uses information on the people selected for the experiment. These data are collected by the administration or through interviews that are part of the experiment.

The timing of such analyses may differ widely. Two specific situations are ideal for its use: when a new programme is going to be introduced and when a programme is going to be

<sup>8</sup> This means that the unemployed person dropped out of the survey, but it is considered as a labour market state.

<sup>9</sup> They make the distinction between four types of programmes: qualification, adaptation or apprenticeship contracts, public interests jobs, courses for preparation to the working life and other courses.

stopped. Some experimental evaluations have also been carried out while the programme was implemented (see Box 3-3).

### *3.1.2.2. Strengths and weaknesses*

If these experiments are correctly done, they are supposed to solve the biases usually found in quasi-experimental studies. However, they do not always solve them and can also have problems of their own.

A specific issue is ethical: it may be difficult to justify refusing help to some people who are supposed to be in need of it in the first place. This is a particularly important problem if one considers that it may lead the administrators implementing the programme to be very reluctant to participate in such an experiment. Evaluators may find it difficult to obtain cooperation from them and may have to make trade-off which can in turn affect the efficiency of the evaluation. The implementation of such an evaluation can therefore be easier in cases where a new programme is tested or where a programme is going to be stopped anyway.

Given that the experiment has to be advertised in some way, several specific biases can emerge: randomisation bias, Hawthorne effect, and disruption bias. The first bias is related to the effect that the announcement of the experiment has on the behaviour of agents. They may choose not to apply to the programme because they may dislike the idea of participating in an experiment. Such an effect may create a significant change in the composition of the group of applicants. The treatment group may not be representative of the usual applicants so that the results of the evaluation cannot be generalized. The second bias (Hawthorne effect) also concerns the behaviour of the participants: if they know they are being studied, they may work harder, and in effect the experiment itself changes the outcome. The final bias (disruption) is in part linked to the ethical problem of such evaluation, in other words, administrators do not behave as usual, whether they are overly enthusiastic or against the experiment. For example, they may select people particularly suitable for the programme (in particular in the “pre-selection” stage), so that the success rate of the participant is artificially high.

This type of evaluation is time consuming (and therefore requires a lot of money) in design and provision of information to administrators (especially if they are reluctant to carry it out). This problem can, however, also be found in the construction of surveys and in other types of evaluation.

Experimental analysis also suffers from substitution bias. It comes from the fact that administrators can voluntarily offer substitute programmes to the people who have been refused the trial programme.

Several examples of these programmes are provided by Björklund and Regnér (1996), one of which is presented in Box 3-3.

This type of analysis is very popular in the United States and some researchers do not hesitate to advocate it as the unique accurate method of evaluation at the microeconomic level (see for example Friedlander et al. 1997). However, it has its shortcomings and relies on assumptions just as quasi-experimental analysis does (Heckman, 1996).

**Box 3-3. The Job Training Partnership Act (JTPA), Björklund and Regnér, 1996, United States**

The programme was introduced in 1982 and administered by Service Delivery Areas (SDA) which were designing the mix of services that would be supplied. The experiment took place across sixteen of these SDAs; people were randomly assigned to the treatment group and to the control group (which counted one third of the people taking part in the experiment). The evaluators found a positive earnings effect for females, a slightly positive effect for males and a negative effect for young people.

The difficulties encountered during the design of this evaluation are very informative. When asked, 90 per cent of the SDAs refused to participate in the experiment. The main reason for refusal was ethical, the administrators also feared that the assignment to the control group would decrease further enrolment in programmes. The problem was that they had excess supply of services at the time of the experiment. The experiment could have also affected the performance measure (monitoring) which is very important in the USA. The evaluators had to consequently spend time and money on informing them, and it was still difficult to convince them. In the end, evaluators had to make some trade off: the SDAs which took part in the experiment were given compensation and the period during which the control group was refused help was decreased from 30 to 18 months.

This shows how difficult it is to implement experiments, the possible reaction of administrators (from whom agreement must be obtained), and also that the timing and more generally the context of the experiment should be taken into account.

## 3.2. Aggregate impact analysis

### 3.2.1. Definition

This type of analysis takes place at a more aggregated level. It is one of the alternatives available to researchers if they want to include the effects of the programme on the rest of the economy (in particular the displacement and substitution effect<sup>10</sup>). Two types of analysis can be used, each corresponding to a different timing. First, the researchers can use a macroeconomic model and macroeconomic data to assess what has happened in the economy and estimate the effects of particular employment policies (Scarpetta, 1996). Second, they may use macroeconomic models to simulate what would happen in the economy if the programme was implemented (DARES, 1997).

In order to evaluate the effect of the programme, the evaluators must choose an underlying theory of the functioning of the labour market. Evaluation but also active labour market policy itself depend on this choice. For example, unemployment may be believed to be the efficient outcome of economic activity; i.e. markets are efficient and unemployment is voluntary. In these theories, government intervention (including ALMPs) is deemed useless or even undesirable. On the contrary, if the underlying theory depicts unemployment as the product of market failures, then government intervention is justified (Snower, 1996).

<sup>10</sup> The other types of analyses estimate directly these effects. Note that here the term “substitution effect” is used for a slightly different phenomenon: the substitution occurs between subsidized and unsubsidized workers within the same firm.

Layard and Nickell (1986) developed a popular model used in aggregate impact analysis. This model is explained in the appendix. Examples of how the Layard and Nickell model has been used are in Box 3-4 and Box 3-5.

**Box 3-4. The Role of Labour Market Policies and Institutional Settings on Unemployment: A Cross-country Study (Scarpetta, 1996)**

In this study, active labour market policies are only one of the factors determining the level of structural unemployment and the speed of labour market adjustment. Other factors include other policy variables (unemployment benefits, employment protection legislation, non-wage labour costs), cyclical factors (the percentage difference between actual and long-run trend output), institutional factors (unions and the wage bargaining system, exposure to trade as a proxy for product market competition) and other factors (real interest rates and the terms of trade). The variable indicating ALMPs consists of the expenditures for ALMPs per unemployed person relative to GDP per capita.

Scarpetta uses the factors described in the previous paragraph to explain the variations of four different variables: the unemployment rate, the youth unemployment rate, the long-term unemployment rate and the non-employment rate. He has data on 17 OECD countries over the period 1983-1993. He finds a negative relationship between the unemployment rate and his measure of ALMPs and between the non-employment rate and the same measure. ALMPs do not have a statistically significant effect on the two other variables (youth and long-term unemployment rates).

He assumes that using a constant for his measure of ALMPs spending (the average over the period) corrects the simultaneity bias. His results concerning the effect of ALMPs on non-employment suggest that ALMPs may have a positive effect on labour force participation. A remaining bias, which is not corrected for in the statistical analysis but is taken into account in the conclusion, comes from the fact that some participants in the programmes might be excluded from the claimant count of the unemployed even though they are looking for work.

### **3.2.2. Strengths and weaknesses**

The main advantage of such an analysis is that it gives the net effect of labour market policies. It often does not give details about the size of the displacement and substitution effects, but studies of outflow rates for certain groups (when such data are available) may identify effects on target groups as well as other groups of people who may be affected (Bellman and Jackman, 1996a). It does not generally account for the dead-weight loss, although some authors have been able to identify it. For example, Schmid et al. (1996) found that the construction sector disproportionately used job subsidies at times when the activity on its market was low.

One of the weaknesses of this technique is that the adjustments subsequent to the implementation of a labour market programme may take time. For example, a short-term increase in unemployment may be observed, while the long-term effect may be favourable. The evaluators must therefore take this into account (and make assumptions about the period of adjustment). One can distinguish here between a conjunctural analysis and a longer-term macro-analysis. The former approach consists of studying the direct impacts of the ALMPs, for example those which have occurred in the past year. The longer-term approach is

interested in the induced effects on the economy as a whole and on the impact of the financing of the ALMPs (DARES, 1997).

Calmfors (1994) also identifies several problems that should be kept in mind when interpreting the results. A weakness common to a lot of empirical studies is that they do not take into account the increase in labour force participation induced by the ALMPs. The effect of ALMPs may also depend on the level of unemployment in the economy: there are theoretical arguments which suggest that ALMPs are more effective with higher unemployment. Finally, there is a simultaneity bias, which is due to the fact that the relationship between programmes' expenditures and unemployment goes both ways: the level of unemployment may change because of changes in government expenditures, but the opposite may also be true.

According to Calmfors (1994) and others, the empirical evidence provided by aggregate impact analysis is still rather limited. Most of the studies concern aggregate wage setting, while some are interested in the effect on employment (or unemployment).

Despite these limitations, we find aggregate impact analysis to be a good way of evaluating programmes. However, in order to have a complete understanding of the effects of a particular programme, it has to be used in conjunction with other techniques.

**Box 3-5. Macroeconomic effects of ALMPs (Bellman and Jackman, 1996b, OECD)**

They use a panel (pooled time-series and cross-section) and examine the determinants of the unemployment rate as well as the effects of policies on some structural economic variables through which ALMPs may affect overall employment and unemployment. They use data on 17 OECD countries over the period 1975-1993.

They investigate the effects of several ALMPs on several labour market variables. The ALMPs (expressed as expenditure per unemployed person) are the following: provision of public employment services, training programmes for unemployed adults and those at risk, subsidies for regular employment in the private sector and direct job creation in the public or non-profit sector (the first three variables are also added together to create a variable measuring the overall expenditure directed towards improving labour market efficiency).

They estimate ALMPs' effect on the following variables: unemployment rate, incidence of long-term unemployment, growth rate of employment, labour force participation rates (of men and women) and two indicators of wage dispersion (the ratio of the ninth and fifth decile and the ratio of the first and fifth decile). The closer these ratios are to one, the smaller the earnings inequality is. They control for the influence of passive labour market policies and institutional arrangements by introducing the following variables: the replacement ratio, the duration of unemployment benefits, the degree of centralisation of wage bargaining, the degree of institutional sclerosis, union density and a variable indicating whether the country allows temporary lay-off.

The approach is similar to that of Scarpetta (1996) except that they use different variables. The results obtained here are not very informative and are hard to interpret without supplementary information. The authors are able to conclude that "the means of ALMPs are far too small to combat the persistently high level of unemployment in Western Europe."

### 3.3. Cost benefit analysis

### **3.3.1. Definition**

This quantitative technique counts, in money terms, the costs and benefits of a particular programme. It compares two alternative courses of events, which must be clearly defined, and evaluates the programme's impact as the difference between the two. The alternative can be what happens under another project, but it generally represents what happens if no policy is carried out at all. The benefits include the net impacts of the programme on the participants (compared to the non-participants) as well as other benefits arising for the rest of the economy (for example, a decrease in crime), the costs include the spending that the programme involves. This technique is generally used as an ex-ante analysis of projects which are under deliberation, but it can also be carried out ex-post (Delander and Niklasson, 1996).

The evaluators have to determine which impacts should be estimated, then how they should be valued and aggregated. The evaluators will calculate the net benefit: the weighted sum of all the individual gains and losses. Generally, the final choice will include losses for some people, but the principle is to implement the programme which is in the interest of the public. The evaluation is based on the "willingness to pay" of the individuals affected. The principle corresponds to an opportunity cost analysis: a favourable impact is evaluated by the maximum sum of money that the beneficiaries would be willing to pay to have it, and an unfavourable effect is evaluated by the minimum sum of money that the sufferers would be willing to accept as compensation for putting up with it (Sugden and Williams, 1978).

One of the main interests of this technique is that it gives a systematic way of approaching policy evaluation. Its main requirements are to clearly identify the primary purpose of the programme and to consider all possible impacts of the programme (as well as its external effects) without restricting them to the quantifiable impacts. The target group (or/and area) of the project has to also be clearly defined (Nas, 1996).

Delander and Niklasson (1996) advocate that a cost benefit analysis of an ALMP should include attempts to estimate the sign, the order of magnitude and the timing of the production impacts, of the income distribution impact (in particular, the gains for the target group) and of the impacts on public revenues and expenditures. It should also include an analysis of the existence and relative importance of other impacts and calculations of present values. Box 3-6 and Box 3-7 show how cost benefit analysis can vary in terms of its scope.



**Box 3-6. Cost-benefit analysis in the JTPA study (Orr et al., 1995, United States)**

This programme has already been mentioned in Box 3-3. The evaluators of the study provide an interesting example of a cost-benefit analysis of active labour market measures.

The costs and benefits are measured as the changes from what would have occurred if the person had not been enrolled in the programme. The costs are the incremental resources used by employment and training services as well as any adverse effects on the people participating in the programme or other people. The benefits include the impacts on the earnings of the programme participants. The net benefits consist of the algebraic sum of all programme's benefits and costs. These benefits are assessed from the point of view of the participant, the rest of society (i.e. non-participants) and society as a whole (i.e. participants and non-participants). For example, a wage subsidy is considered as a benefit for the participants, but as a cost to the rest of society (as in effect, it pays the subsidy). In the end, for society as a whole, the effect is null (it is the sum of the effect on participants and of the effect on the rest of society). The table below shows which impacts the evaluators expected the programme to have, as well as the direction of these impacts. Not all of these impacts can be measured, and in effect, the evaluators evaluated the impacts in dollars only for the starred items in the table.<sup>11</sup>

	Participants	Rest of Society	Society as a whole
Earnings gain (minus wage subsidy) *	+	0	+
Wage subsidy *	+	-	0
Fringe benefits	+	0	+
Increased taxes on earnings *	-	+	0
Welfare benefit reduction *	-	+	0
Incremental training costs *	+/-	-	-
Reduced leisure time and home production	-	0	-
Increased work related expenses	-	0	-
Reduced criminal activity	+	+	+
Psychological benefits of increased employment	+	+	+

They estimated these impacts for different target groups (adult males and females, young males and females), as well as for each different service provided under the JTPA (classroom training, job subsidy, and other services). They used data obtained through a background information form, follow-up surveys and official sources. They chose an impact period of 30 months. They also carried out some sensitivity analysis of their results, changing their measurement techniques; the results were unchanged.

They found beneficial effects for adult men and women, while the rest of society suffered costs and society as a whole gained. They however found negative effects for young men and women, costing a lot to the rest of society and society as a whole.

All services provided in the programme brought positive benefits to adult men and women and to society as a whole (except classroom training for women which represented a cost to society as a whole) and negative benefits to the rest of society. All services cost money to the participants, the rest of society and society as a whole when young people were considered (except classroom training which brought a positive benefit to young female participants).

<sup>11</sup> They also carried out an evaluation using an estimate of the value of the fringe benefits, but the results were not changed.

The approach is consistent with what is advocated in the theoretical literature. Compared to what could have been done, this cost-benefit analysis is however very limited in the number of impacts it is able to evaluate.

### 3.3.2. *Strengths and weaknesses*

As for evaluation and policy choices in general, the use of cost benefit analysis depends on the institutional structure and the corresponding incentives given to the decision-makers. This technique is quite ambitious in its scope. It is indeed very difficult in practice to have a clear alternative to the programme studied. Analysts may have incomplete information on the direct and indirect effects of the programme. Moreover, certain effects, such as psychological effects, may be difficult to measure in money terms. Several methods exist to estimate the value of those effects: the method of revealed preferences is based on the observation of the behaviours of individuals with similar characteristics to the programme's participants (in particular, the way they choose between two alternatives). The stated preferences or contingent valuation method uses the answers given by respondents to questions involving choices between carefully selected alternatives. The value of these effects may also be based on the evaluators' experience, beliefs or observation. In any case, such methods or equivalent methods should be applied to cost benefit analysis of ALMPs, so that a maximum number of effects can be accounted for.

Another issue is that the impacts of the programme do not occur all at the same time. A period of evaluation has to be chosen. In general, only short-term effects are considered. However, when one wants to evaluate the long-term effects, costs and benefits have to be considered over a number of years. In this case, one has to calculate the value in the present year of all future benefits and costs (i.e. their present value). The future value of benefits is discounted by a rate similar to interest rates. There is a large literature on the choice of this discount rate, and people have often used several reasonable rates. The other way to evaluate long-term effects is to calculate the rate of return; it corresponds to the discount rate for which the present value of the benefits equals the present value of the costs. The rate of return of a particular programme can then be compared to those of other programmes.

The results provided by a cost benefit analysis should always be considered in a wider perspective. Even if the cost benefit ratio (value of benefits divided by value of costs) is a good indicator of the effectiveness of a programme, it should not be considered in isolation. In particular, the unmeasured effects should be taken into consideration, and the evaluators should also explain how these impacts occur. The cost benefit ratio may not be appropriate for all programmes, so that other methods of evaluation may have to be used (Ryan and Grubb, 1998).

This type of analysis enables the evaluators to study not only the net benefit for society, but also for different groups of people, i.e. the redistributive effect of the programme. Its results, which are generally concise and clear, can be a very useful tool in the formation of future policy. Even though a perfect evaluation may never be reached, cost benefit analysis may decrease the risk that decisions are made inefficiently. It will also contribute to making the decision-makers accountable to other decision-makers and ultimately to the voters.

**Box 3-7. Cost benefit evaluation of the Job Corps  
(Delander and Niklasson, 1996, USA)**

Job Corps is a programme which provides vocational training skills, basic education and health care in residential centres. It aims at improving the employability of economically disadvantaged youths. The evaluation was carried out by Long et al. (1981).

The estimates of the programme's effects are based on the comparison of the outcomes of the participants and a matched group of non-participants. The data come from an initial and two follow-up interviews (the last taking place on average 18 months after participation). Longer-term benefits were estimated by extrapolating the interview data under the hypothesis that their magnitude would decline over time. Prices are evaluated at the time of the entrance in the programme and the discount rate used is 5 per cent per year. Each benefit and cost is evaluated from the point of view of society, the Corps members and the rest of society (transfers between the Corps members and the rest of society cancel each other out as far as society is concerned). Not all the effects identified are measured. This is similar to what is described in Box 3-6; however, a monetary value is given to many more effects. For example, they evaluate the reduced criminal justice costs, the reduced personal injury and property damage, the reduced value of stolen property, the reduced treatment costs for drug and alcohol use (all these effects coming from the fact that the youths are off the streets).

They found, in their benchmark estimates, a large positive benefit (in money terms) for the participants and a comparatively small cost to the rest of society, so that the net effect for society as a whole is largely positive. They carried out some sensitivity analysis but the effects of the programme seem to stay positive.

This study is a very good example of what can be achieved with cost benefit analysis. It is necessary, in terms of policy recommendation, to outline the various redistribution effects in addition to the cost benefit ratio.

### **3.4. Qualitative analysis**

#### **3.4.1. Definition**

This type of evaluation takes into account the opinion of the agents concerned by the programme: beneficiaries, administrators implementing the programme and firms. The data is obtained through surveys and interviews. These data are used to estimate the effects that cannot be quantitatively measured. Sometimes qualitative analysis simply consists of an assessment made by experts in the field concerned, based on their experience. Other times, journalistic accounts of programmes' results may also be considered as qualitative analyses.

#### **3.4.2. Strengths and weaknesses**

This is a useful way of judging the quality of the ALMPs (see Box 3-8 for an example). It may indeed be impossible to measure certain of their aspects, for example the quality of the training provided or the efficiency of the administrative staff. This technique is also used to evaluate the dead-weight loss, substitution and displacement effects. For instance, the evaluators ask the employers whether they would have hired the participants in the absence of the programme, and they ask the employees whether they would have accepted the job without the subsidy.

This technique suffers from the usual problems with survey data. The respondents may interpret the questions asked in various ways. Surveys also suffer from non-response and attrition of the sample of interviewees.

Qualitative analysis is probably carried out in a less systematic way than the quantitative techniques that have been studied so far. It is often based on small samples, which are not representative of the entire population, sometimes even simply on a single individual case study.

This technique is therefore less rigorous than the others and the accuracy of its conclusions cannot generally be verified, but it probably gives a good idea of factors which are relevant for a complete evaluation and which cannot be measured with precision. In order to obtain more objective views, the evaluators have to use carefully designed and simple multiple choice questions, such as “what is the destination of the individual after participation in the programme?”, “would you [the firm] have created this new job if this programme did not exist?”, etc. Questionnaires can also include crosschecks, asking similar questions in different ways to the same person, and asking different agents similar questions.

Because of their low cost, qualitative studies are particularly useful in cases where little money is devoted to evaluation. Combined with quantitative data obtained through monitoring (which are always collected), they can provide a useful low cost evaluation.

**Box 3-8: A qualitative study of the “Investors in People” programme ( Rix et al., 1994, United-Kingdom)**

The Investors in People programme aims at promoting training and development in UK companies. The state offers official recognition once certain standards have been achieved.

Rix et al. provided a small-scale qualitative study on the effects of the programme on employers. Their study was based on 25 employers and 19 groups of employees. The research was concerned with identifying issues of importance to employers and employees as well as their perceptions and feelings about their experiences of the programme. The companies studied formed three groups: (i) those formally recognized, (ii) those formally committed to working towards the National Standards, (iii) those uninvolved. The sample was too small for the results to have any statistical reliability.

Their overall assessment of the programme was positive, although some possible improvements in the implementation of the programme were identified. For example, it seemed that the companies having achieved the standards had already a great interest in the training and development of their staff, and the programme only provided the “final touch” to their own policy.

Overall, despite its limitations this qualitative study gave useful insights in the implementation of the programme. The authors were able to draw some conclusions which were considered to be useful for future implementation and policy.

### **3.5. Lessons for determining a framework for evaluation**

These techniques have been developed for several decades and most involve advanced econometric knowledge. There is indeed a large literature reviewing the different aspects of these techniques, in particular the ways to improve the estimates. Their use therefore requires a certain expertise and they are best carried out by specialists in the field. Moreover, no single technique comes out as ‘the best’ evaluation tool. Each of them has its advantages and

drawbacks depending on the type of programmes to be evaluated. They are complementary, not alternatives and should be used in combination in order to have a complete evaluation.

Several issues, common to all techniques, have been identified. First, the collection of data is a crucial issue in quantitative policy evaluation. The scope of the evaluation will depend on the quality and quantity of the available data. Second, <sup>24</sup>cette partie devrait être non-technical it seems crucial to have some sensitivity analysis of the results, especially if these results are going to be used in policy making, in particular in decisions to stop particular programmes (Heckman and Smith, 1997). This can be seen in experimental and quasi-experimental analysis for which the results can vary greatly. Sensitivity analysis consists of checking whether the results are affected by changes in the assumptions taken by the analyst. For example, in the case of cost benefit analysis, the analyst may vary the discount rate used. Finally, all evaluation methods require assumptions, the most important step in solving evaluation problems is to evaluate the plausibility of those assumptions in each particular context (Heckman et al. 1996).

The determination of the evaluation methodology, i.e. the choice of the combination of techniques to be used, will depend in part on the institutional structure and the administration of the country. They set the budget, which in turn determines the scope of the evaluation. This is a particular issue considering that nowadays governments look for ways to decrease their budget deficits. As far as ALMPs are concerned, the evaluation part is the easiest one to cut, as it does not affect the agents directly. The authorities also set the aims, which lead the choice of techniques, according to their own interests. They also are the ultimate users of the evaluation's results, and it has been shown that they do not often take them into account.

## **4. A framework for evaluation**

This section aims at giving a framework to evaluate active labour market policies in view of what is advocated in theory and what is feasible in reality. It gives a suggestion for a “practical guide to evaluation”, which will have to be adapted to the particular target group considered.

### **4.1. A target-oriented approach**

Section 4 uses the evaluation ‘philosophy’ of Schmid et al (1996) who advocate a target-oriented evaluation as opposed to evaluating single policies in isolation.

It has been argued that although evaluation of specific programmes is useful in itself, it has several drawbacks: in particular, it does not take into account the interaction of the particular programme with its environment (institutions and other economic policies) and it does not enable comparisons between the effects of various programmes. A target-oriented approach enables the evaluators to take these into account. The idea is to compare<sup>12</sup> the outcomes of different labour market programmes targeted at particular groups and to evaluate the programme within its environment (institutions, values, other policies). The ‘target groups’

<sup>12</sup> The techniques used to evaluate different programmes should be comparable enough to be able to choose between them.

include people who suffer from specific difficulties in the labour market and who are generally entitled to receive additional help in looking for a job (such as long-term unemployed, young people, or old people). Another reason to choose a target-oriented approach is the following. The aims of the governments in carrying out ALMPs have evolved over time, and will probably evolve over the years to come. In the seventies and beginning of the eighties, they were primarily interested in finding new ways of decreasing unemployment. Since the mid-eighties, with the aggravation of the situation of certain categories of people, they have searched for policies which avoid ‘social exclusion’ (Mouriaux, 1995). By taking a target group approach (with a broad definition of target group), one can design an evaluation approach which does not depend on the governments’ aims and can be reproduced in the future. In practice however, few evaluations have yet adopted such an approach (Meager and Evans, 1998).

#### **4.2. Policy formation stage**

To improve the usefulness of evaluation reports, there must be a feedback between the evaluators and policy makers. This implies that the evaluators present clear recommendations and that the government takes them into account. The overall scope of evaluations remains dependent on the political will and the institutional framework. Transparency of government’s goals and practices should ensure more reliable evaluation results.

Evaluation can be best carried out if it is decided at this stage (the policy formation stage). This would, among other things, enable the evaluators to obtain the data which will be useful to their analysis. In addition, the evaluation strategy should follow set principles: for example, we have suggested here the use of a target-oriented approach. As well as ensuring a systematic evaluation of all policies and continuity in policy formation,<sup>13</sup> this would also enable policies to be better co-ordinated and consistent with each other.

A collaboration between the agents involved (the political, administrative and research worlds) can facilitate the choice of combination of techniques. It can also allow the divergent interests of these three agents to be taken into account. The short-term view of politicians has to be reconciled with the necessity for researchers to study longer-term effects. Box 4-1 gives a “real life” example of the French evaluation experience.

<sup>13</sup> The empirical evidence gathered on the countries’ practices as regard ALMPs (Gautié et al.) shows that policies are changed every few years. This creates an unstable environment. Several authors (Ryan and Grubb, 1998) have recognized that there is a need for more stability and continuity in policy formation, as well as in programmes’ evaluation. Continuity in policy formation may be difficult to achieve given that policy-makers often have a short-term view. One can introduce continuity in the evaluation process by creating a benchmark against which any policy would need to be evaluated. This benchmark would consist of a set of principles which form a skeleton for evaluation.

**Box 4-1. The evaluation of ALMPs in practice: the French example (DARES, 1996).**

The evaluation of ALMPs in France started at the end of the seventies. An evaluation programme was launched in 1977 in response to an OECD project aiming at measuring the efficiency of employment subsidy programmes. However, there was no systematic development of evaluation analysis in subsequent years. Although long-term macroeconomic evaluations have been carried out and have enabled the evaluators to identify the overall effects of policies on employment and unemployment, there has been no equivalent of such analyses at the microeconomic level.

From the mid-eighties, all large ALMP have been regularly evaluated, using administrative data, surveys of the participants (unemployed and firms), etc. These studies have all been specific to particular programmes and carried out only at one point in time (for example, a few months after the start of the programme's implementation). This approach to evaluation has enabled the evaluators to obtain results which were not restricted to estimating only some quantitative impacts chosen in view of the aims of the policy. Their results were also used in policy making. However, there was no global approach to evaluation, so that the programmes could not be compared to each other. Moreover, these evaluations always looked at gross outcomes, i.e. they did not use control groups. The evaluations were also restricted by the budget they were given and the interests of the various governments.

Quasi-experimental analyses have only been used recently in France, and experimental analyses have never been used, because of technical as well as ethical issues. The recent quasi-experimental analyses suffer from the same problems as their English or US equivalent: their results are often contradictory, and no definite conclusion can be drawn. They have also focused on employment outcome, forgetting other impacts of policies.

Recent qualitative studies have pointed out the need to take into account the opinion of the participants in programmes as well as the employers' view. There seems to be a lack of understanding between the policy makers and the agents (firms and workers), so that more and more policies are implemented, giving fiscal advantages to firms, while, overall, the latter have a negative view of these policies.

The experience of policy evaluation in France provides a good example of all the other factors on which the success or failure of a policy depends and which should be taken into account in an evaluation. These are the following: the financial means given to the administrators of the policy (considered to be insufficient in France); the interactions between policies (French policies are considered to be complex and to include numerous measures and agents, as well as being often modified); the practical side of the implementation (the administrator must face the local situation of the labour market, and must conciliate divergent requirements from the various agents: employers, administration, unemployed, etc.).

Moreover, the evaluation can start here: the degree of feedback can be evaluated. The evaluators can see whether the policies are consistent with previous evaluations' findings. At the same time, the evaluators can check whether the policy-makers are responsive to the different needs of individuals and to the changing economic conditions. They can see whether the needs of the target population have been identified correctly (the evaluators themselves should identify these needs). All the programme's possible effects need to be identified, even those which will not be studied in the evaluation.

### **4.3. Policy implementation stage**

The implementation of the policy is a crucial stage, and plays an important part in the success or the failure of the policy. Several factors need to be looked at in addition to the general requirements of the target-oriented approach.

The first is the quality of the implementation itself. The success of the policy depends on the quality of the services provided to the participants. The dimension of quality is not often included in evaluations. The usefulness of such an evaluation depends on the particular measure considered, some (such as employment services) may for example aim at reaching as many people as possible - so that quantity is more important than quality - while others (such as training) may be effective only when their quality is recognized. At the same time, it may be useful to evaluate whether the administrators are competent, although this may be difficult to achieve.

An evaluation at the implementation stage can make use of the information concerning the local labour market. Evaluation cannot be solely done at the local level, as there would be no coherence in it (each local office being primarily concerned with its own welfare). The information gathered by local offices must therefore be centralized and used in an overall evaluation. Such information is very relevant to evaluation. First, the programme must answer local problems. For example, while an increase in the places in nursery education may be thought to enhance mothers' participation in the labour market (in view of national evidence) a local study could reveal completely different needs, such as a need for help with elder persons' care. Second, in order for the policy to be successful, it has to fit into the local system. Local administrators can facilitate the implementation by organising co-ordination with the agents at the local level (firms, associations, local authorities, etc.). Third, local administrators of the programme have also to deal with the participants themselves. In that way, evaluation also consists of checking whether the policy is appropriate to the motivation, interests and capacities of the target groups. The individuals' behaviour may not always be consistent with what is expected by the administration. At this level, one needs to check whether the services reach the target group and how they are used by the beneficiaries. For instance, it may be useful to determine the degree of abuse of the policy (Mouriaux, 1995). Finally, an important dimension to implementation is whether it is cost-effective. One has to evaluate whether the budget is efficiently allocated. This can be done in comparing the achievements of two different local agencies with the same budget (and facing similar labour markets).

The evaluation takes on the findings of the monitoring exercise: it uses quantitative data such as the number of participants, the structural composition of participants, the dropout rate, etc. In addition, it aims at explaining why the implementation is successful or not. The latter can be done by interviewing the agents and gathering qualitative data.

### **4.4. Impact stages**

These stages are generally the most studied in policy evaluation. The aim of impact evaluation is to determine whether the participation in a programme has improved the situation of the unemployed person or the situation of a group of unemployed persons. The impacts are various, but evaluations have focused on the employment and earnings effects. Such evaluations are now considered to be insufficient by most authors (see for example DARES (1996) or Mouriaux (1995)). The latter insists on the need to include a qualitative analysis of the programmes. More generally, the social impacts of the programmes are deemed important in a context of high and persistent unemployment (Kraft, 1998).



A way of considering all relevant effects may be to list all the possible impacts of the programme. These impacts can be classified in many different ways, here four broad groups are defined: the direct impacts on the participants (employment and earnings outcomes, improvement in well-being), the social impacts (improvement in social equity, improvement in social environment through a reduction in crime for example), the fiscal impacts (changes in tax revenue and government spending), the impacts on the labour market (improvement in matching between jobs and workers, change in the unemployment rate, change in workers productivity, change in other groups' unemployment). Good evaluations will also include an analysis of the timing of these effects, as well as data requirements, although the latter would have been done at an earlier stage. Interviews with the unemployed (and more generally the agents involved in the programme) are another important source of data. This is already done in practice, and the questionnaires that are usually used could be extended to include a larger number of qualitative questions. Administrative data (obtained through monitoring) can be very useful in quantitative analyses.

The impact evaluation may be more or less broad (i.e. address more or less issues) depending on its future use. Most of current evaluations are partial; they address only specific questions so that their findings cannot be generalized. Although specific questions can be selected in view of the interests of policy-makers, a complete evaluation should address a broad range of issues. Box 4-3 gives an example of questions which can be addressed in an impact evaluation of youth labour market policies.

The statistical techniques need to be chosen in view of the purposes and means of the evaluation: aggregate impact and qualitative analyses seem to be the cheapest, the quantitative analyses seem to require a lot of data gathering which may be expensive, but are useful to have precise estimates, cost benefit analysis may be too ambitious but it gives a general philosophy which is interesting to follow. Box 4-2 gives two examples of combinations of techniques. The weight given to each of these techniques may vary greatly according to the country's government and institutional structure. The idea is to improve the current practice of evaluation and to highlight the fact that if it is not organized and recognized as part of the policy process, it is simply a waste of time and money.

**Box 4-2.            Examples of combinations of techniques**

- (vi) In cases in which the State wants to introduce a completely new programme or stop one, experimental (microeconomic) analysis can be carried out. In addition to using administrative data, the evaluators have to design questionnaires and interview participants and non-participants several times in order to obtain a longitudinal dataset. Moreover, a macroeconomic simulation can be used for the aggregate impact analysis of a new programme or macroeconomic data can be used to estimate the impacts that an on-going programme has had on the labour market. Some kind of cost benefit analysis can also be carried out. The evaluators should list and explain all the possible impacts of the measure, even if all of these are not measured. Qualitative analysis should be included by adding some qualitative questions to the interviews, and by interviewing employers and programme administrators as well as the unemployed.
- (ii) In cases in which the State wants to evaluate an on-going programme: it is best to use a quasi-experimental (microeconomic) analysis, together with an ex-post aggregate impact analysis. Again, some kind of cost benefit analysis may be useful, and a qualitative analysis should be included.

**Box 4-3. An example; impact evaluation of policies targeted at youths.**

The justification for government intervention on the youth labour market comes from several observations (O'Higgins, 1997). In recent years, characterized by high unemployment levels in the countries considered in this paper, the youth unemployment rate has been higher than the adult rate and the proportion of youths experiencing long-term unemployment has been higher than what is commonly believed. It is reasonable to say that experiences of unemployment early in the career may impair future labour market outcome (in terms of employment and earnings) for the affected individual. There has been evidence that "unemployment leads to more unemployment". Moreover, youth unemployment is associated with other social problems such as crime and drug use. These are additional reasons for the government to intervene. The aims of the policies targeted at the youths have therefore been to solve the above mentioned issues. In effect, they have been designed to smooth the transition between school and work. One can consequently ask several questions which need to be answered about the performance of the policies. (Examples of ALMPs targeted at the youths and their evaluation can be found in Deakin (1996), O'Higgins (1994), Long et al. (1981), and others.)

- (i) Direct impacts on the participants (as opposed to the non-participants):
  - Do they have a greater probability of finding regular employment?
  - What was the duration of the first job they found after completion of the programme?
  - Do they have greater earnings than they would have had?
  - Do they have a lower probability of coming back to unemployment?
- (ii) Additional (external) impacts:
  - What are the sizes of the dead weight, substitution and displacement effects (with respect to all other unemployed or to other target groups)?
  - Are there positive externalities for the rest of society (such as reduced crime or reduced drug consumption)?
- (iii) Macroeconomic impacts:
  - Has the unemployment rate (employment) of the youths comparatively decreased (increased)?
  - What is the effect on the size of the labour force?
  - What is the effect on total unemployment?
  - What is the effect on average youth's earnings?
- (iv) Items to consider in cost benefit analysis (examples can be found in Deakin (1996) and Delander and Niklasson (1996)):
  - Change in output, tax revenue due to programme (both after and during the programme).
  - Change in welfare transfers.
  - Change in criminal activity.
  - Change in drug/alcohol use.
  - Change in use of other ALMPs.
  - Change in well-being and income distribution.
  - Programme operating expenditures.

## 5. Conclusion

This paper has provided a review of the literature concerned with the evaluation of active labour market policies, with a focus on what is being done in the European Union and the United States. It has shown that evaluation has not been systematically carried out (in Europe in particular), and that in a country like the United States, which has developed the evaluation techniques very far, its use in policy making has been rare and controversial. Given the usefulness of active labour market policies and the accountability of the government to the public, the paper has considered the use of policy evaluation to be essential.

It has therefore provided a guide to how policy evaluation should be designed. There are practical limitations to ALMPs evaluations, so that a ‘complete’ evaluation may never be feasible. However, the following general principles can be recalled and outlined:

(i) **A target group approach seems appropriate.** Evaluating single programmes may be of more limited utility. First, these are generally implemented within a general strategy of ALMPs, and it is more interesting to evaluate this strategy. Second, policies interact with each other, sometimes being complementary, sometimes substitutes to each other, the usefulness of a programme therefore also depends on its interactions with other policies. Third, the results of single programme evaluations cannot be compared. In a target-oriented approach, evaluators are able to solve these issues.

(ii) **The determination of evaluation strategies seems preferable early in the policy making process, in co-operation with the relevant agents.** All parties (policy-makers and evaluators) should work in collaboration to determine an evaluation strategy early in the policy making process. In order to be fully successful, evaluation can be introduced as a systematic rule so that there is some continuity even when there is a change of government. For example, one can use a common framework, such as the one described in this paper. This framework must be adaptable to new situations. It would enable the evaluators to compare among policies. This may not always be feasible in practice, but the idea is to build in evaluation as part of the design and implementation stages rather than as an afterthought.

(iii) **The usefulness of evaluation studies is enhanced when their results are used in policy making.** It has been sometimes argued that long-term evaluations may be out of date (and therefore useless) once they are completed. Although politicians generally require rapid information on the success of a programme, a long-term evaluation is considered here to be very useful in future policy, even if new (and different) policies are implemented. It indeed gives an idea of how the different agents of the labour market respond to particular aspects of policies. Moreover, as long as the target group approach is taken, the evaluation may help to shape new policies, simply in providing the information needed to improve previous programmes. As with social science research in general, the use of evaluation results in policy making may take different forms. This question however goes beyond the scope of this paper.

(iv) **Monitoring** can be used to rapidly give indications of the success of a programme through comparisons of the performance of the agencies implementing it. As, among other things, it provides valuable information and data, an efficient monitoring is useful for evaluations.

(v) **“Transparency” is necessary to sell new policies to the taxpayers.** Evaluation results need to be trusted by the public. Internal as well as external evaluations can be improved by enabling the public and outside researchers to check and challenge these.

(vi) **The evaluation should not only consider the impacts of the programme** but also its implementation and formation. It seems to be forgotten that the economic agents are the ultimate determinants of the success or failure of a policy. If firms do not understand measures they may not use them. Unemployed people may abuse the weaknesses of certain programmes. Others may simply waste their time on compulsory programmes. Local administrators may be under-qualified or local agencies may be understaffed to be able to apply a programme. More generally, the administrators and the users are a substantial source of information to explain the achievements of a programme.

(vii) **As many of the impacts as possible need to be evaluated.** Employment and earnings effects give only a partial view of the impacts of the programmes on the labour market and on agents. Other aspects, in particular, social and equity aspects of labour market policies are relevant to society's well being. Moreover, in order to put the results in perspective, the evaluators need to state the time period they are considering, and to mention the possible longer-term effects.

(viii) **A combination of evaluation techniques should be used.** It has been shown that no single technique can provide a tool for a complete evaluation, only a combination of techniques will. We have seen that, even in countries where there is no evaluation, there is a kind of control (through monitoring) on what is going on, and that some countries rely on gross outcomes. It therefore makes sense to spend a little bit more and do better evaluations either with a control group or with a qualitative analysis trying to determine what would have happen otherwise.

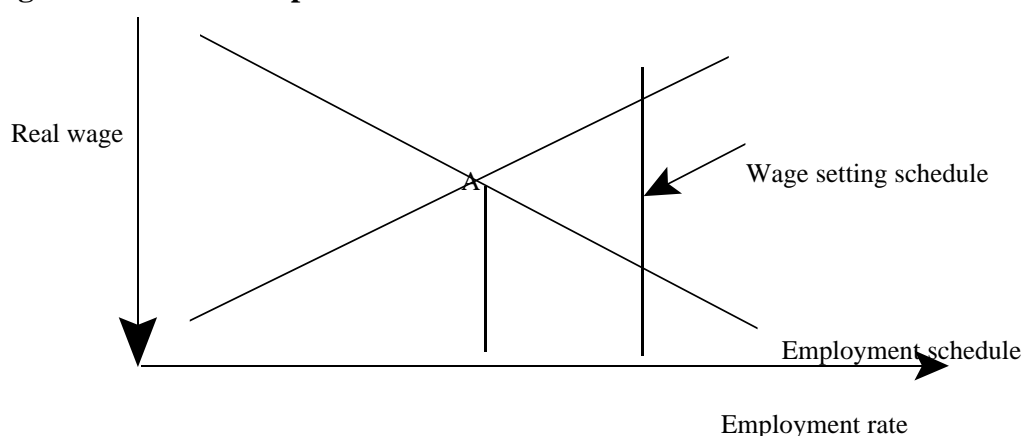
(ix) **In order to improve the quality of the results, data from various sources can be used.** For macroeconomic data, it has been argued that using different sources for the same series may provide more accurate results. At the microeconomic level, different type of data should be gathered, in particular, qualitative data have been found to be a good complement to quantitative data.

Finally, although this paper has focused on examples of evaluations carried out in Europe and the United States, the lessons learned in these countries can be useful for developing or transitional countries which come across similar economic difficulties. The general framework and ideas developed for policy evaluation can be applied in those countries, within the constraints of their institutions (government will, financial means, etc.).

## Appendix: The Layard-Nickell model

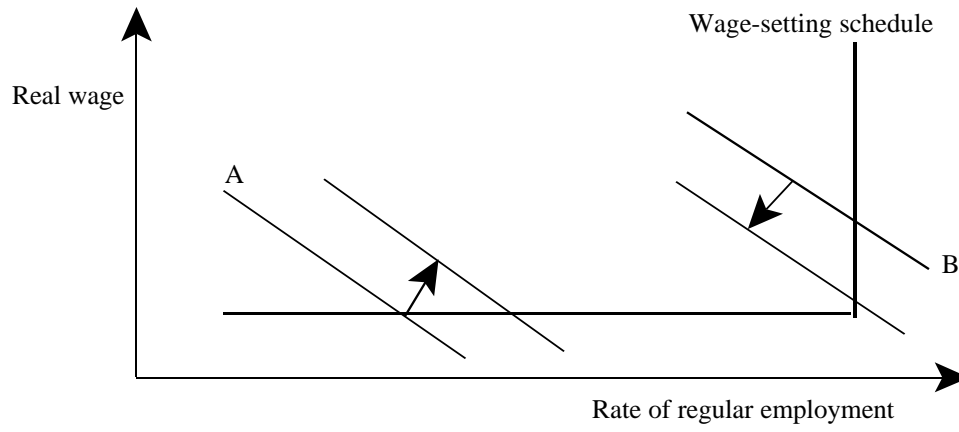
The simple model (see figure 1) consists of a wage-setting schedule and a price-setting (or employment) schedule. The former shows a positive relation between real wage and employment, i.e. how higher aggregate employment causes pressure for higher real wages. The employment schedule represents a negative relationship between real wages and employment: it shows the number of workers the firm would employ at each level of the real wage. The labour force is taken as given. The equilibrium employment and real wage are at the intersection (A) of the two curves. The difference between labour force and the employment level determined by the intersection of the employment and the wage-setting schedules represents unemployment. This model can be refined (for example by introducing the Beveridge curve, which shows the relationship between unemployment and vacancies) in order to analyse the effects of ALMPs, which have been reviewed at the beginning of this paper (see Calmfors, 1994).

**Figure 1. The simple model**



We show in figure 2 a stylised wage-setting schedule: it is horizontal when there is unemployment and vertical at full employment. The low-productivity and high-productivity sectors are represented by employment schedules A and B respectively. A transfer of labour from the low-productivity sector to the high-productivity sector (for example through training) is represented by the shifts in the employment schedules: in the low-productivity sector the demand for labour increases as a proportion of the labour force, while in the other sector it decreases. We can see that after the reallocation has taken place, employment rate is higher in the low-productivity sector and has not changed in the high-productivity sector. A larger share of the labour force is in the high employment rate sector, so that overall, the employment rate has increased.

**Figure 2. Reallocation of labour**



## Bibliography

- Auer P. (1997), "Activation of Labour Market Policy in the European Union - Comparative Summary", Employment Observatory, SYSDÉM, Trends No. 28, European Commission.
- Auer P. and Kruppe T. (1996), "Monitoring of Labour Market Policy in the EU Member States", Wissenschaftszentrum Berlin für Sozialforschung, Discussion Paper FS I 96-202.
- Bellman L. and Jackman R. (1996a), "Aggregate Impact Analysis", chapter 5 in Schmid G., O'Reilly J. and Schömann K. (eds), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar.
- Bellman L. and Jackman R. (1996b), "The Impact of Labour Market Policy on Wages, Employment and Labour Market Mismatch", chapter 24 in Schmid G., O'Reilly J. and Schömann K. (eds), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar.
- Björklund A. and Regnér H. (1996), "Experimental Evaluation of European Labour Market Policy", chapter 3 in Schmid G., O'Reilly J. and Schömann K. (eds), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar.
- Bonar Blalock A. ed. (1990), *Evaluating Social Programs at the State and Local Level. The JTPA Evaluation Design Project*, W. E. Upjohn Institute for Employment Research, Kalamazoo, Michigan.
- Bonnal L., Fougère D. and Sérandon A. (1997), "Evaluating the Impact of French Employment Policies on Individual Labour Market Histories", *Review of Economic Studies*, 64 (4), pp. 683-713.
- Calmfors L. (1994), "Active Labour Market Policy and Unemployment - A Framework for the Analysis of Crucial Design Features", *OECD Economic Studies No. 22*, Paris.
- Calmfors L. and Lang H. (1993), "Macroeconomic effects of active labour market programmes - the basic theory", Seminar paper No. 541, Institute for International Economic Studies, Stockholm University.
- Calmfors L. and Nymoén R. (1990), "Real wage adjustment and employment policies in the Nordic countries", *Economic Policy*, No. 11.
- Cette G., Cuneo P., Eyssartier D. and Gautié J. (1996), "Coût du travail et emploi des jeunes", in Revue de l'OFCE, *Observations et diagnostics économiques*, janvier, OFCE, Paris.
- DARES (1996), *40 ans de politique de l'emploi*, La Documentation Française, Paris.
- DARES (1997), *La Politique de l'Emploi*, Repères, La Découverte, Paris.
- Deakin B. (1996), *The Youth Labour Market in Britain. The role of intervention*, Cambridge University Press.
- Delander L. and Niklasson H. (1996), "Cost-benefit Analysis", chapter 6 in Schmid G., O'Reilly J. and Schömann K. (eds), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar.
- Fay R. (1996), "Labour market and Social Policy", Occasional Papers No. 18, OECD, Paris.
- Freeman, R (1998), "War of the Models: Which Labour Market Institutions for the 21<sup>st</sup> Century?", *Labour Economics*, 5, pp. 1-24.
- Friedlander D., Greenberg D. and Robins P. (1997), "Evaluating Government Training Programs for the Economically Disadvantaged", *Journal of Economic Literature*, 35, pp. 1809-1855.

- Gautié J., Gazier B., Silvera R. et al. (1994), *Les Subventions à l'Emploi. Analyses et expériences européennes*, Document Travail et Emploi, La Documentation Française, Paris.
- Gregg P. and Wadsworth J. (1996) "It Takes Two: Employment Polarisation in the OECD", CEP Discussion Paper No. 304, London.
- Heckman J., Ichimura H. and Todd P. (1998), "Matching As An Econometric Evaluation Estimator", *Review of Economic Studies*, 65, pp. 261-294.
- Heckman J. and Smith J. (1996), "Experimental and Nonexperimental Evaluation", chapter 2 in Schmid G., O'Reilly J. and Schömann K. (eds), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar.
- Heckman J. and Smith J. (1997), "The Sensitivity Analysis of Experimental Impact Estimates: Evidence from the National JTPA Study", National Bureau of Economic Research, Working Paper No. 6105.
- Kraft K. (1998), "An Evaluation of Active and Passive Labour Market Policy", *Applied Economics*, vol. 30, no. 6, pp.783-793.
- Layard R. and Nickell S. (1986), "Unemployment in Britain", *Economica*, pp. S121-S169.
- Layard R., Jackman R. and Nickell S. (1991), *Unemployment*, Oxford University Press.
- Long D., Mallar C. and Thornton C. (1981), "Evaluating the benefits and costs of the Job Corps", *Journal of Policy Analysis and Management*, 1 (1), pp. 55-76.
- Meager N. and Evans C. (1998), "The evaluation of active labour market measures for the long-term unemployed", Employment and Training Papers No.16, ILO, Geneva.
- Mouriaux M-F. (1995), "Croyance, expérience, évaluation: l'évolution de l'aide aux chômeurs créateurs d'entreprise", in *Les politiques publiques d'emploi et leurs acteurs*, Cahiers du Centre d'Etudes de l'Emploi 34, Presses universitaires de France, Paris.
- Nas T. (1996), *Cost-benefit Analysis. Theory and Application*, SAGE, London.
- O'Higgins N. (1994), "YTS, employment, and sample selection bias", *Oxford Economic Papers*, 46, pp. 605-628.
- O'Higgins N. (1997), "The challenge of youth unemployment", Employment and Training Papers No.7, ILO, Geneva.
- Orr L., Bloom H., Bell S., Doolittle F., Lin W. and Cave G. (1995), *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study*, The Urban Institute Press, Washington, D.C.
- Payne J., Lissenburg S., White M. and Payne C. (1996), "Employment Training and Employment Action. An evaluation by the matched comparison method", Department for Education and Employment, Research Series No. 74, London.
- Rix A., Parkinson R. and Gaunt R. (1994), "Investors in People. A qualitative study of employers", Employment Department, Research Series No. 21, London
- Rossi P. and Freeman H. (1989), *Evaluation: A systematic Approach*, Fourth Edition, Sage Publications, London.
- Ryan P. and Grubb W. (1998), "Plain talk on the field of dreams: the roles of evaluation for vocational education and training", ILO, Geneva.



- Sachs J. (1987), "High unemployment in Europe: diagnosis and policy implications", in Siven C-H. (ed.), *Unemployment in Europe*, Timbro, Stockholm.
- Scarpetta S. (1996), "Assessing the Role of Labour Market Policies and Institutional Settings on Unemployment: A Cross-country Study", *OECD Economic Studies No. 26*, pp. 43-98.
- Schmid G. (1996), "Process Evaluation: Policy Formation and Implementation", chapter 7 in Schmid G., O'Reilly J. and Schömann K. (eds), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar.
- Schmid G., O'Reilly J. and Schömann K. (1996), "Theory and Methodology of Labour Market Policy and Evaluation: An Introduction", chapter 1 in Schmid G., O'Reilly J. and Schömann K. (eds), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar.
- Schmid G. and Reissert B. (1988), "Do Institutions Make a Difference? Financing Systems of Labour Market Policy", *Journal of Public Economics*, vol. 8, no. 2, pp. 125-149.
- Schmid G. and Reissert B. (1991), "On the Institutional Conditions of Effective Labour Market Policies", in *Beyond Keynesianism. The Socio-economics of Production and Full Employment*, Matzner E. and Streeck W. eds., Edward Elgar.
- Schöman K. (1996), "Longitudinal Designs in Evaluation Studies", chapter 4 in Schmid G., O'Reilly J. and Schömann K. (eds), *International Handbook of Labour Market Policy and Evaluation*, Edward Elgar.
- Shapiro C. and Stiglitz J. (1984), "Equilibrium unemployment as a worker discipline device", *American Economic Review*, Vol. 74.
- Snowder D. and de la Dehesa G. (1996), *Unemployment Policy. Government Options for the Labour Market*, Cambridge University Press.
- Sugden R. and Williams A. (1978), *The principles of practical cost-benefit analysis*, Oxford University Press.
- Warburton W. (1996), "What Went Wrong in the CETA Evaluations?", *Canadian Journal of economics*, 29, pp. S105-108.
- Weiss C. (1979), "The many meanings of research utilization", *Public Administration Review*, 39 (5), pp. 426-431.

Copyright © International Labour Organization 1999

Publications of the International Labour Office enjoy copyright under Protocol 2 of the Universal Copyright Convention. Nevertheless, short excerpts from them may be reproduced without authorization, on condition that the source is indicated. For rights or reproduction, or translation, application should be made to the ILO Publications Bureau (Rights and Permissions), International Labour Office, CH-1211 Geneva 22, Switzerland. The International Labour Office welcomes such applications.

Libraries, institutions and other users registered in the United Kingdom with the Copyright Licensing Agency, 90 Tottenham Court road, London W1P 9HE (Fax:+ 44 171 436 3986), in the United States with the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923 (Fax:+ 1 508 750 4470), or in other countries with associated Reproduction Rights Organizations, may make photocopies in accordance with the licences issued to them for this purpose.

---

---

The designations employed in ILO publications, which are in conformity with United Nations practice, and the presentation of material therein do not imply the expression of any opinion whatsoever on the part of the International Labour Office concerning the legal status of any country, area or territory or of its authorities, or concerning the delimitation of its frontiers.

The responsibility for opinions expressed in signed articles, studies and other contributions rests solely with their authors, and publication does not constitute an endorsement by the International Labour Office of the opinions expressed in them.

Reference to names of firms and commercial products and processes does not imply their endorsement by the International Labour Office, and any failure to mention a particular firm, commercial product or process is not a sign of disapproval.

ILO publications can be obtained through major booksellers or ILO local offices in many countries, or direct from ILO Publications, International Labour Office, CH-1211 Geneva 22, Switzerland. Catalogues or lists of new publications are available free of charge from the above address.

---