Benchmarking carrots and sticks: developing a model for the evaluation of work-based employment programs

Castonguay, J.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
4. Performance indicators of work-based employment programs

The previous chapter has served to place work-based employment programs within its context, to give it a functional definition, and to present the three dimensions at the basis of its intervention strategy. It is now clear that those programs are based on a multitude of hypothesis on which type of interventions are needed to assist the unemployed in returning to the labour market. These will be used throughout this chapter in order to choose the performance indicators for the benchmark model which is being developed.

This chapter will thus follow the policy chain which was presented as the basis for the social benchmark model in chapter 2. Section 3.1 had concluded that the benchmark of work-based employment program should be a best-practice benchmark, taking into account the whole policy-chain when measuring its performance. This will allow not only to determine the efficiency and effectiveness of the program, but also to open the black-box of policy implementation and evaluate the effect of different processes on the impact of the programs. The question to be answered in this chapter is thus: which performance indicators are required for benchmarking work-based employment program on the basis of the policy-chain? The input, process, output, impact and external factors indicators will thus each be presented, providing an explanation for their relevance in the model, as well as the way the indicators should be measured and ranked.

4.1. Input indicators

As mentioned in chapter 2, the input refers to elements that are initially present for the policy-makers in order to implement a program. In other words, input are the foundation upon which a program will be built. The first and most obvious of these indicators is the one referring to the objectives of the program. The objectives of most labour market programs are very multi-faceted and shaped by many different assumptions, hypothesis and paradigms. Often, these are not black-and-white propositions; it can even happen that different aspects of the objectives contradict each other. Benchmarking the objectives of a program will thus require a deep understanding of the political context of the program in order to better understand how these are being formulated.
By differentiating main objectives from secondary objectives and understanding the origin of objectives, a closer measure of effectiveness and efficiency can than be attained. This is because both measures of effectiveness and efficiency are relative to the objectives of the program. Also, a distinction can be made between the explicit objectives of a program, being those which are part of the official literature on the program and which are used openly in the political arena, and the more implicit objectives which are not made official but still are to be attained by the program. This means that the source of information will include official publications from the governments, and laws and regulations will be an important source for this. Indeed, some countries also mention in their legislation what a particular law is meant to achieve. Since laws form the major building bloc of social programs, objectives found in the legislation can be said to be the most formal and explicit objectives of a program. Another distinction which can be made is between the levels of the objectives. Objectives can have to do with the effect of the program on the individuals themselves (micro-level objectives), or they can have to do with the effect on the economy as a whole (macro-level objectives). An example of this difference is that while increasing the skill-level of the unemployed is a micro-level objective, decreasing the size of the caseload is considered a macro-level objective. All these different types of distinctions (main or secondary, implicit or explicit, micro or macro level) should be taken into account in the qualitative analysis of this input indicator in order to give more depth to the measuring of the different types of objectives in each program.

Concerning the benchmarking of the different objectives as such, a distinction is made between objectives which have to do with the impact of the program, and objectives which have to do with the output of the program. As already explained in chapter 2, the output of the program are the activities which take place, such as training activities, while the impact of the program is the actual effect these activities have on the participants. Concerning the impact-objectives, since work-based employment programs are part of a more general strategy to assist the unemployed back to the labour market, the major impact to be reached is the return to the labour market of the participants. Therefore, the first sub-indicator for the objectives is the extent to which the return to the labour market forms the goal of the program.

Besides focusing on the return-to-work of the participants, it is also possible that the focus of the objectives of the program may be more towards the output. The output of work-based employment programs will be discussed later on in section 4.3, but three broader categories of output-oriented objectives can be defined for this section. These three
categories are in fact directly linked with the intervention-strategy as defined in the previous chapter. As discussed earlier, work-based employment programs attempt to increase the rate of exit-to-work through three different mechanisms: increasing the ability to work, increasing the willingness to work, and increasing the access to work. The output-oriented objective of work-based employment programs can therefore be defined as aiming at increasing either of the ability, the willingness and the access to work. More concretely, this means that, first, raising skill levels can be an important objective of the program. The objective of raising skills level and employability clearly refers to the output of the program concerning the amount of training offered, either through formal training component of the program or though the work-activities. Second, increasing the willingness-to-work through the creation of a (mutual) obligation within the benefit scheme can also be an objective of the programs. In order to achieve this, work-activities can be required in order to establish an obligation for the claimant to “give something back for what he/she receives”. But willingness-to-work can also have to do with increasing the earnings prospect of the participants when they do find a job, and can thus also refer to activities which increase future productivity, such as training. Lastly, the program can have for objective the improvement of the access to work of its participants. In terms of activities (or output) this can be related to interventions intending to improve signalling towards potential employers as well as interventions aiming to prevent claimants from becoming socially excluded. Summing up, four different objectives can thus be distinguished, that is to say: 1) Return to the labour market, 2) increase employability, 3) create mutual obligations, and 4) increase access to work.

Each of these four objectives will be measured separately as one or more of them can be present in the program. As all other indicators, this indicator will be measured on a scale of 1 to 5, where 1 will indicate that the objective is not at all present in the program, and 5 will be the country with the highest focus on this objective within the benchmark. This is because this benchmark uses a best-practice benchmark methodology, where the highest scoring country always receives the highest score. The other countries will be judged against this standard, scoring them between the spectrum of 1 and 5 according to their relative position to the other countries with respect to the importance and strength of each objective. Since the focus on the return to work will be the most important indicator against which the results of the programs will be judged, this sub-indicator of the objectives will be used to give a score to the countries in the benchmark. The other sub-indicators relating to the
output will be used in the qualitative analysis of the effectiveness of the programs in chapter 11.

The second input indicator to be included in the benchmark is the **target group** of the program. Active labour market policy was typically of two kinds, universal programs such as the one found in Sweden in the 90s and targeted programs such as the ones found in many countries in the rest of Europe, such as the UK (Dorstal, 2008, p.26). The groups targeted by labour market programs may be very diverse and vary greatly in specificity. For example, a program may be targeted a long-term unemployed lone mothers with low levels of education, which selects the participants based on many different characteristics. Otherwise, programs can be targeted at all long-term unemployed, no matter their other characteristics. Work-based employment programs, here defined as one specific targeted ALMP program, are thus also targeted at some specific groups. Some target groups may be defined by law, as part of the general eligibility criteria for a benefit. For example, legislation can state that young people under the age of 25 and who have been unemployed for more than one year are required to take part in the work-based employment program in order to be able to receive a benefit. Since this target group will be defined by law, it will be very difficult for those implementing the program to divert from this definition. On the other hand, target groups can be very vaguely defined through official government documents, and left more to be defined on a case-by-case basis. In such a model, much freedom is given to those implementing the program in defining the actual target group of the program. Measuring target groups will thus require an analysis of juridical elements in the programs, as well as information from the level of the implementation.

Within the programs in this benchmark, two groups can be differentiated with respect to work-based employment programs: the young unemployed and the long term unemployed. That is not to say that participants cannot be characterised according to other groups, based on gender, ethnicity, family status, education level, and more. This distinction is made between these two target groups because these are specifically being targeted by work-based employment programs. Special distinctions are usually not made to make other categorisations, such as migrants, or women, the specific target of the provision. This will also mean that often, data will not be available on the presence of these different categories of claimants in the benefit. Nevertheless, where relevant, other possible target groups will be added to the discussion. It should be noted again that the fact that the only target group of all American programs is lone mothers meant that it was excluded from this benchmark.
The precise definitions of these target groups vary greatly between countries, but some standards have been set by international organisations such as Eurostat and the OECD. The standard definition for long-term unemployment adopted by Eurostat and the ILO counts as long-term unemployed those who have claimed a benefit for 12 months or more. Youth unemployment is usually seen as occurring between the ages of 15 to 24 inclusively, but the operational definition is likely to vary from program to program (O’Higgins, 1997, p.63). As explained above, the input indicators will measure the extent to which the “foundations” behind the process of the program can be expected to lead to good results. This means that the larger the input indicator, the more this indicator should assist in achieving good results. When looking at the target group, this means that the indicator should measure the extent to which the target group is expected to have good chances to find a job through the program.

With respect to young unemployed, a large part of the target group being young would affect the rate of return to the labour market in a positive way. Even though youth unemployment rates are usually higher than total population unemployment rates, it is considered that within similar type of employment programs, the younger participants would have a higher rate of exit in times of economic growth. As explained by O’Higgins (1997, p. 69), aggregate demands has a much stronger influence on youth unemployment rates, meaning that it will decrease much stronger than total unemployment in times of economic boom. Since the year 2006 (the year in which the data will be collected in part 2) was a year of economic growth for most industrial countries, this would mean a proportionally higher level of return to work for young people in most countries. Furthermore, according to the OECD (2003, p. 210), international comparison points towards an important difference between younger and older unemployed, whereas for the young unemployed there is little link between benefit dependency and employment. Additionally, two country-specific studies on work-based employment programs, Sol et. al. (2008) for the Netherlands, and Bolvig et. al. (2003) in Ochel (2007, p.86) for Denmark, showed that young participants were more successful in returning to work.

1 See Eurostat Key indicators on EU policy – Structural indicators – Social Cohesion: Long term unemployment rate; Methodological Summary; and, KLIM 5th Edition Ch. 10: Long-term unemployment, p.1. The OECD gives two separate values for Long-Term unemployment, that is to say, unemployment at 6 months and more as well as unemployment at 12 months and more (see Employment Outlook 2007, p. 267).
The opposite is true for long-term unemployed, who thus have much smaller chances to find a job since they have been claiming unemployment benefits for a long time. The correlation between length of claim and lower chances to exit unemployment has been discussed by many authors, such as Jackman and Layard (1991); Van den Berg and Van Ours (1994); Machin and Manning (1998); and many more. Clearly, depreciation of skills, motivation to work, social exclusion as well as discrimination all play a role in decreasing the chance the long-term unemployed do find unemployment. At the opposite of targeting long-term unemployed are programs which are rather focused on early-intervention. These programs are targeted at only the new claimants who are not yet inside the benefit system. This would make the program much more of a “gate-keeper” type of program where the focus is much more the prevention of entry, as well as on the prevention of loss of willingness of capacity and willingness to work if entry does occur.

In sum, these two groups will be measured separately and then averaged to provide a measurement of overall target group chance of finding a job. Hence, the country where only the youth unemployed are the targeted will receive a score of 5 and the one where only those unemployed for less than one year (thus not long-term unemployed) are targeted will receive a score of 5. Both scores will then be averaged to provide for a single ranking based on the distance from the labour market of the overall target group in each program.

The third component of the input indicators are the financial resources and the incentives brought about by the distribution mechanism of these financial resources. What is being measured here is not the way the budget is allocated within the program, for example the resources allocated to private providers. This will be part of the process indicators. Rather, what is being looked at is the general budget allocated by the government to the program, and the extent to which the mechanisms which allocate this budget to the government organisation responsible for the implementation also create some incentives for this organisation to reach good results. This is measured since these incentives also measure to what extent the input in place can guarantee that the objectives will be met.

When comparing budgets of programs in different countries, many cautionary elements must be taken into account. Since it is crucial that the budget numbers all measure the same part of the program, only the budget allocated to the employment services related to the work-based employment program should be taken into account. This includes any element of training, job search assistance, and any other type of
provisions, but excludes benefit expenditure and other expenditures related to allocations for the participants. Benefits, allocations and wage subsidies should not be included since two separate indicators will measure first the generosity of the benefit and second the generosity of the program’s allocation. This indicator of the financial resources available for the program will thus only refer to the financial constraints of the program, with regards to designing its main provisions. A country with a large budget will thus be able to offer more services (in quantitative and qualitative terms) than a program with a small budget. A large budget is thus expected to lead to better results. Furthermore, when comparing government spending on ALMP in percent of Gross Domestic Product, authors such as Classen and Clegg (2003, p.371) have suggested dividing this percentage by the number of unemployed, allowing to compare the spending indicators on the basis of their intensity. Indeed, a large budget, even if related to the GDP or other indicator of the size of the total government budget, would be more an indication of the number of participants than of the generous financing of the program. Since program size (in term of coverage) will be part of the output indicator, the indicator here should clearly compare the amount of budget which is available to the program. Hence, this indicator regarding the financial resources should divide the annual budget for the employment program by the number of participants in that year. Also, needless to say that a common currency should be used in order to convert those individualised budget values to comparable units. Clearly, the largest indicator will receive a score of 5, and the lowest a score of 1, while the other countries will be ranked relative to the figure of these two cases.

The second aspect of this indicator of the financial resources allocated to the program measures its incentive mechanism. What is being compared is the extent to which the government body responsible for the implementation of the program is being given a financial incentive towards reducing caseloads as much as possible. For example, block grants where any surpluses or shortcomings are usually covered if needed present much less incentives than when the surpluses can be kept or must be added upon by the implementing government body. Similarly, clear target agreements between the central government and the body implementing the program which would link results with future budgets would represent a strong financial incentive. The country with the strongest incentive through financial allocation will thus receive a score of 5, since this would lead to increased efforts towards reaching good results. In the absence of any incentives, the country will score a 1. In order to make-up this indicator, the score for the size of the budget will be averaged with this score on incentives.
The next input element to be discussed is the division of responsibilities within the government agencies delivering different aspects of the program. Much reform has happened in the last decades in the way governments deliver any type of services, and employment services have not been left out. These aspects of the New Public Management mode of governance which have increasingly influenced public service delivery have been introduced in chapter 1. As was explained, in terms of the division of the responsibilities for the delivery of work-based employment programs, two different trends can be distinguished. First, with the rise of New Public Management, bureaucracies were criticised for the scattered nature of service delivery, which was the product of the specialised nature of public work (Finn, 2000). On one side, one body would take care of determining eligibility and paying out benefits, and on the other side, another body would provide employment services such as job search assistance and training programs. However, this situation was deemed inefficient for many reasons, such as the need for two different intake procedures as well as the lack of communication between these two public bodies. For example, the case manager responsible for the delivery of employment services would often not be aware of the claim history and specific eligibility agreements of a participant, and vice versa. Many countries thus brought changes to this segregated structure since it was believed that the individual needs would be better met by combining functions which are closely related to each other. New Public Management thus meant the integration and coordination of services which were to be delivered by integrated agencies, often called “one-stop-shops”. Since this reform towards the concentration of services within one body is expected to increase efficiency and effectiveness, the more those services are concentrated within one body, the higher the country will score in the benchmark.

Secondly, a trend of decentralisation has also been taking place. This was brought about by inefficiencies caused by the lack of local adaptability of programs within many countries where programs were uniformly designed and implemented from a central government level. As pointed out by Finn (2000), there came a need for more flexible institutional arrangements, especially considering the nature of local labour markets which are only partially connected and highly differentiated in most countries. Finn thus points out the three ways in which this flexibility can be achieved, proposing that this can be done through first, the geographical targeting of national measures, second, the devolution of responsibility to regional or local authority, and third, through the creation of local partnerships between the national implementing body and local actors. Furthermore, Finn (2000) summarizes the benefit of this decentralisation and flexibilisation. The author shows how involving
local actors allows for pilots and experiments to be undertaken, makes it possible to adapt the program to local reality, how this increases inter-linkage (synergies) through local partnerships, finally how new actors increase scale of delivery and create a broader local network. Consequently, by allowing decentralised provision of employment services, the results of the work-based employment programs are expected to be better. Hence, the countries with the most decentralised delivery system will thus receive a score of 5, while the country with the least decentralisation will score a 1. Both score for concentration and decentralisation will be averaged to provide for a single indicator of the efficiency of the division of responsibilities with respect to the provision of the work-based employment program.

These first input indicators were drawn from hypothesis about the influence of the macro-level incentives on the results of the program. On the other side, the next three input indicators will now discuss the influence of three important inputs on the individual chances to return to the labour market. These micro-level inputs are the unemployment benefit in which the program will take place, the activation conditions present in this benefit scheme as part of the general eligibility requirements, and the sanctioning mechanisms in place which enforce these activation conditions in the benefit scheme.

The first micro-input indicator relates to the nature of the benefit scheme itself in terms of type and benefit level. This indicator is one of those indicators which are found in the legislation of each country. The analysis of the laws and regulations concerning social security benefits will thus be central to measuring this performance indicator. Special attention will be given to the basic structure of the benefit payment, such as whether it is made up of different component which are added up to form one total benefit. By using laws and regulations directly in this indicator, a better understanding of how benefits levels are determined for each individual will be possible. Indeed, only using benefit levels available in international database such as the OECD does not explain how benefits may vary from one individual to the next within the benefit scheme itself. Using laws and regulations directly as a source of information will thus show what average rates are not able to express.

First, a qualitative discussion should explain the type of benefit scheme in which the program will take place. These will be divided into three basic category which are related to its basic eligibility criteria: social assistance.

---

2 These are three standard categories, as used amongst many others by the OECD (see OECD, Benefits and Wages 2007, p.16-23-26).
(means-tested) and unemployment insurance (contribution-based) and unemployment assistance (means-tested, for those whose eligibility to unemployment has ended, thus also contribution based). This discussion on the type of benefit will not result in a ranking since there is no clear hypothesis relating the type of benefit and the expected results to be achieved by the program. In fact, the major differences between the nature of the three categories will be encompassed into the indicator of the generosity of the benefit, which is the second aspect to be included in this indicator. Also, since the benefit level also forms the means-test for most schemes, this will also be an indicator of the income level of those who are eligible for this benefit.

This measurement of benefit generosity will be relevant for all work-based employment programs, even though some programs might pay-out a wage to participants meaning they are not part of the benefit scheme anymore. This is because the participants will usually enter the program thought their claim for the benefit, even though this claim will automatically be cancelled by an immediate participation in the program. Even in this case, the incentive structure and eligibility criterion of the benefit scheme will have an influence on the participants entering the program. For example, a generous benefit payment might actually be the reason why a potential participant will initially decide to claim a benefit.

The level of the benefit should – similarly to the indicator of financial resources – be measured using similar units. Also, since standard of living can vary from one country to the next, crude measures of benefit levels might not give a clear idea of the real generosity of a benefit. For this reason unemployment benefits are usually expressed in terms of replacement rates, where the benefit are measured relative to the wages earned by average workers (see OECD, 2007b, annex A.1 p.60). However, measuring replacement rates relative to average wages is much more relevant for unemployment insurance since these benefits are meant to replace previous income from work (Nelson, 2008). The concern here is not with the benefit generosity with regards to previous income from work, but from the perfective of activation. It would therefore be more appropriate to calculate a measure of the benefit generosity by relating them to the statutory minimum wage in each country. This will give a clearer indication of the extent to which the benefit is more or less attractive than finding a regular job for which minimum wage is being offered. An other possibility would be to use the Average Effective Tax Rate as calculated by the OECD which give a measure of the incentives to enter the labour market when one is already unemployed (see OECD, 2007b, p. 107). However, this measure of the “inactivity trap” in each country is only calculated for unemployment insurance benefit, while as
will be seen all programs in the benchmark are part of social assistance or unemployment assistance. In conclusion, since an indication of the relative incentive structure of each benefit system is needed, the best available indicator will be the benefit level as a percentage of a full-time work week at the net minimum wage salary. However, as Nelson (2008, p.9) warns, social assistance benefits are often set in order to meet the basic needs of a whole family, whereas minimum wages are individually set. This should be taken into account when calculating this indicator, and comparing single households without children might be the most comparable approach to this measurement.

Furthermore, since some benefits are given on a fortnight basis while others on a monthly basis, it is best to bring all amounts to a measure of all benefits received within one year, relative to net yearly income from a full-time minimum wage job. In addition, any supplements for which all or most claimants have a right to receive should be added to this amount, such as holiday supplements. It might also be possible to deduct earnings from work in order to still claim the full benefit even though income might not be null. In the presence of the possibility to deduct some of the earnings form work, a proportion of these earnings will not reduce the benefit level. This would in fact raise the total amount of income which is available to benefit claimants. In addition, these deductions can be in the form of a taper, in which higher levels of earnings from work have an increasing deduction rate up to the point where the income earned is fully deducted from the benefit to be received. Nevertheless, it is most likely that participation in the work-based employment program will not be mandatory for those who do work part-time while claiming the benefit, so this should be taken into account in deciding whether this earnings deduction should be taken into account.

The ranking of this indicator on the 1 to 5 scale of the benchmark is a lot thornier than most other indicators. This is because conflicting evidence exists on the effect of benefit generosity on the inflow and the outflow of the unemployed in the benefit scheme. First, from the perspective of the outflow out of the benefit, there has been considerable research done on the influence of benefit level and length of claims, showing that higher benefit levels lead to longer claims (see Lalive, Van Ours and Zweimüller (2006); Fortin, Lacroix and Drolet (2004); and Roed and Zhang (2003) for recent evidence). The influence of the benefit level and the incentive to enter the labour market has already been introduced in the previous chapter. There, it was shown how the intersection between the productivity rate and the utility for work and leisure determine the reservation wage of the person. In the case this reservation wage is lower than the benefit level, the person will be better off claiming the benefit.
Clearly, the higher the benefit level, the more chance the reservation wage of the person will be found lower, which means that the incentive structure needs to be altered if activation is to take place. This analysis would imply that the rate of return to work will be lower for higher benefit levels in the work-based employment programs. As a result, the country with the least generous benefit will be expected to have the largest exit-to-work ratio and thus will receive a score of 5 in the benchmark. Similarly, the country with the highest benefit level will receive a 1, and the remaining countries will be ranked according to this spread between the benefit levels of those two extremes.

However, from the perspective of the inflow into the benefit, a low benefit level is also more likely to act as a deterrent to claim benefits in the first place. This would thus mean that some of those who would be able to find a job on their own might be more inclined to not claim a benefit. As a result, those with greater barriers to work would make up a larger part of the caseload. Actually, the fact that benefits are lower also means that eligibility is much more targeted at those with very low income, which also can be associated with greater barriers to work. This would lead to lower rates of exit to work in the program. Unfortunately, it would be impossible to measure the extent to which this is true for the participants in the work-based employment programs in the benchmark since productivity measures would be needed for the average participants. Nevertheless, this does not mean that this factor will be completely ignored in the qualitative analysis of this indicator in each country. In the case there are wide differences in benefit levels, the extent to which a very low benefit level is expected to induce larger exit to work will be put in the light of the hypothesis that this would also imply a target group which might have a much lower productivity level. Secondly, the low benefit themselves can be a cause of the delayed return to the labour market because of the social exclusion and financial stress they could bring about. Gallie, Paugam and Jacobs (2003) have indeed showed with data from the European Community Household Panel that poverty makes it more difficult to return to work, leading to a vicious circle in which unemployment and poverty become entangled. This last point can be taken into account by looking at the position of the benefit level with respect to the poverty threshold of the country, where a very low benefit with respect to the poverty line could also imply lower rates of return to work. Similarly, Nelson (2008) calculated the adequacy of social assistance benefits by averaging different levels of benefits in each scheme (varying with family composition) with respect to the average disposable income in the population and found important differences between schemes, and between family types within one scheme. The findings from his calculations should be included in the analysis of this indicator.
for the countries in the benchmark. The cut-off point for severe poverty will be set at less than 40% of median disposable income, where less than 30% would be considered extreme poverty, consistent with categories as used by Behrendt (2002, p.26) when looking at the adequacy of social assistance in three countries. Hence, benefits lower than 40% of the median disposable income would then mean that a lower score would be given for a country whose benefit would actually score highest if only taking into account the evidence on reservation wages and benefit levels as discussed above. This thus implies a U-shaped relation between benefit levels and exit-to-work ratios, where both extremely low benefit levels as well as high benefit levels are assumed to have a negative impact on the results the programs can be expected to achieve.

Concerning this discussion on the benefit level, it should be noted here that within the process-benchmark, an indicator will allow to show whether policy-makers have decided to diverge from this benefit level through either giving a supplement or a wage to the participants in the work-activities. As will be discussed then, this process indicator will be ranked so that more generous rewards will lead to higher rates of return to the labour market. Such a construction will lead to the situation where the “general” benefit level is seen as inversely proportional to the result within the input indicators while the “specific” rewards is seen as directly proportional to the results within the process indicators. This is deemed here to be a realistic representation of the diverging incentive structures which are created by the level of the benefit, as discussed above. In this way, the process is revealing the policy-choices made by those designing the program, within the constraints of the scheme as defined by the input.

As discussed in the previous chapter, work-based employment programs are part of a general trend towards the more pro-active “activation” of the unemployed. Within the general benefit scheme in which the program will take place, activation conditions are therefore present in the legislation which are meant to oblige the unemployed to maximise their efforts towards finding (and keeping) a job. In other words, these activation conditions have the potential to offset some of the negative incentives that arise from the presence of the benefit scheme as discussed earlier (OECD, Employment Outlook 2000, Ch.4, p.129). As mentioned in the previous chapter, these activation conditions actually alter the choice between “work” and “leisure” to become a choice between “work” and “actively searching for work”. As discussed by the OECD (2000, p. 131), three mechanisms will influence the claimant. First, a direct behavioural effect will be noticed, in which the requirements to actively search for work or to participate in employment programs will directly increase the chance to finding work. Second, a disutility effect in which the
requirements to be fulfilled are not seen as much enjoyable for the participant, who will then increase its efforts towards finding a job. And third, an exit effect, where if the disutility effect is so large, the claimant will actually rather quit the benefit system than having to comply with the requirement.

Laws and regulations, and especially the articles referring to benefit eligibility, will thus be the focus of the comparison between the countries. Nevertheless, it is apparent from using legislation in different countries as a primary source of information that eligibility and conditions for the payments of benefits are not a blue print from each others. Amongst others, some countries lay out activation conditions not as a basic eligibility criterion for the benefit, but rather as a behaviour on which a sanction is applicable. The analysis will thus also pay attention to whether the activation conditions are directly written into the eligibility criteria of the benefit, or indirectly through sanction provisions. Moreover, in order to facilitate the comparison of legislation which is written in various forms, activation conditions will be divided amongst four types of requirements. These are, 1) Availability for work and acceptance of work offered, 2) job search requirements, 3) participation in employment programs requirements, and 4) requirements to sign an "activation" contract. Hence, these four requirements will be looked for and inventoried directly from the social security legislation with respect to the unemployment benefit in which the work-based employment program is taking place.

In light of the three effects illustrated above, the stronger these requirements are, the greater the return to work during or after the work-based program is expected to be. Nonetheless, in case the participants in the work-based employment program are actually not claiming the benefit anymore (due to the fact for example that they are paid a regular wage), the actual requirements within the program should replace those found in the scheme in order to make a realistic comparison. Also, as highlighted by the OECD (2000, annex 4A), many differences exist in the definitions within these requirements, such as for example the definition of “suitable work”. These should of course be taken into account when comparing the strength of each requirement, since for example while two countries require that suitable work must be accepted, one country might actually have a much stricter definition than the other. Finally, the actual score for this indicator will be averaged out over the four components, and the country with the strongest obligations with respect to each requirement will receive a score of 5, while a score of 1 will be given to countries where the requirement is not present.
In order to foster the compliance to these activation conditions, **sanctioning legislation** has been put in place within social security laws. These sanctions can be used both as a threat and as a consequence for not following the activation requirements, as discussed above. The effect of sanctions has been investigated by many authors. The large majority of these studies show a positive correlation between sanctions and return to work (Ochel, 2005). Abbring, Van den Berg and Van Ours (2005) found for Dutch unemployment insurance a doubling of the re-employment rates for those who had received a sanction, due to both a lowering of the benefit level as well as the increase in job search in fear of more severe subsequent sanctions. Van den Berg, Van der Klaauw and Van Ours (2004) found similar effects for social assistance recipients in the city of Rotterdam, and also pointed to the fact that this effect was long lasting and felt even after the sanction expired and benefits had returned to normal levels. Svarer (2007) found for Denmark that even moderate sanctions more than doubled the chance of exiting and that the harder the sanction the stronger the effect. Boone, Sadrieh and Van Ours (2004) pushed their analysis further by estimating separate effects for the threat sanctions can pose (a compliance effect, or ex ante) and for the actual effect when the sanction has been imposed (an ex post effect). Their experiment shows that the compliance effect from the mere presence of the sanctioning mechanism (the ex ante effect) are large and even larger than the actual effect cause by receiving a sanction itself. This was also found by Svarer (2007), who showed that those with the most chance to receive a sanction actually experienced the greatest compliance effect. Boon, Frederiksson et al. (2007) even suggest that the gains in increase in return to work stemming from monitoring job search and sanctioning when appropriate will actually be high enough to compensate for the cost of implementing such a monitoring and sanctioning system.

It is thus clear that in this benchmark, the presence of a sanctioning mechanism in the legislation should be associated with better return to work ratios. However, what is less clear is how the severity of the sanction (either in level or in duration) should be expected to correlate with results. Obviously, the threat effect and benefit reduction effects should be important enough so as to make a real difference in search intensity. One can thus assume that very soft sanctions would be less effective than harsher sanctions. However, it is unlikely that there would be a lower limit at which soft sanctions would have no effect at all. This is based on the findings of Van den Berg, et. al. (2004) where a doubling of the re-employment rates was based on sanctions which represented a 20% reduction of the benefit for 1 or 2 months.
On the other side, it is possible to imagine that very harsh sanctions would not always result in better return-to-work ratios. One reason for this is related to the discussion on benefit levels above, where very low benefits (which are in this case the result of harsh sanctions) have a negative impact on the ability of the individuals to find a job. Also, very harsh sanctions can lead to avoidance activities, where the higher the sanction the higher the resources invested in order to avoid it, in such a way that this can divert some resources away from legitimate job search (Frederiksson and Holmlund, 2006). Boon, Frederiksson, et al. (2007) also pointed out that in the presence of imperfect monitoring and risk aversion, maximal sanctions would not result in improvement in welfare. This is because the possibility of very harsh sanctions being given to claimants who do actually comply with the requirements but are the victim of an imperfect monitoring system would largely affect the welfare effect of the system. However, the precise turn off point where sanctions become too harsh to bring about an improvement in exit rates was impossible to measure by the researchers, since much more information is needed about amongst other the costs of implementing the monitoring and sanctioning system. Similarly for this benchmark, while the ranking should take into account the possibility that the country with the harshest sanction would actually have lower return-to-work ratios than for example the country with the second best ranking, this does not need to be the case. In fact, it could be so that the country with the harshest sanctions is still below the point where sanctions become too hard to show improvements. So while the ranking will follow the hypothesis that the more severe the sanction the better the results are expected to be, there will be room left to interpret this ranking in the light of the analysis of Boon, Frederiksson, et al. (2007) as well as the discussion on the adequacy of benefits mentioned earlier.

Concerning the measurement of the severity of the sanctions, some calculations will need to be performed in order to properly compare this indicator. As already shown by the OECD (2000, p. 135), sanctions can vary greatly between countries, not only with respect to the length of the sanction but also in the amount of benefit being withdrawn. In fact, while some sanctions can be short but imply the complete stop of the benefit such as in Denmark, other countries such as Australia rather partially reduce the benefit for a longer period of time (OECD, 2000, p.135). This

---

3 This contrasts with seminal work of Becker (1968) on crime and punishment, where the optimal fine would be the maximal fine. However, his assumptions were that individual are risk neutral and that sanctions can be applied without costs, which is less realistic when looking at sanctioning in unemployment benefits systems.
makes crude comparisons impractical. Also, converting the sanction in monetary units is not useful since the benefit level may vary greatly as discussed earlier.

Table 4.1  Performance indicators for the input of work-based employment programs

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Sub-indicators</th>
<th>Score = 5</th>
<th>Score = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td>Return to work</td>
<td>Strongest focus on return to work</td>
<td>Weakest focus on return to work</td>
</tr>
<tr>
<td></td>
<td>Mutual Obligation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ability to work</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Willingness to work</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Access to work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target group</td>
<td>Youth</td>
<td>Easiest target group</td>
<td>Hardest target group</td>
</tr>
<tr>
<td></td>
<td>Long-term unemployed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finances</td>
<td>Budget per participant</td>
<td>Largest budget and incentives</td>
<td>Smallest budget and incentives</td>
</tr>
<tr>
<td></td>
<td>Incentive structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Division of</td>
<td>Concentration</td>
<td>Strongest orientation towards efficiency in SS system</td>
<td>Weakest orientation towards efficiency in SS system</td>
</tr>
<tr>
<td>responsibilities</td>
<td>Decentralisation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benefit</td>
<td>Type of benefit</td>
<td>Least generous scheme</td>
<td>Most generous scheme</td>
</tr>
<tr>
<td></td>
<td>Benefit level</td>
<td>* lower boundary 40% poverty line</td>
<td></td>
</tr>
<tr>
<td>Activation conditions</td>
<td>Job acceptance</td>
<td>Strongest activation requirements</td>
<td>Weakest activation requirements</td>
</tr>
<tr>
<td></td>
<td>Job search requirements</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participation requirements</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Signing a contract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sanction legislation</td>
<td>% of income loss on</td>
<td>Strongest financial impact</td>
<td>Weakest financial impact</td>
</tr>
<tr>
<td></td>
<td>yearly basis from 1st</td>
<td>* possibility of an upper boundary, defined</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sanction</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What is important here is to measure is the financial impact a sanction will have on the income of claimants. Indeed, severe sanctions with great financial impact would be more likely to compel the claimants to follow the activation requirements as laid-out in the legislation. This indicator will thus be measured by looking at the percentage of income loss on an annual basis as a result of a first sanction. For example, loosing 50% of a benefit for one month would result in a yearly income loss of about 4%. Such a calculation will be done on the basis of the information on sanction
length and sanction levels as found within the legislation. As already mentioned above, the country which will show the greatest financial impact of a sanction will thus receive a score of 5, but some attention will be paid to whether it would be possible that this sanction would be so severe as to actually hinder job search and return to work. The country with the least severe sanctions will receive a score of 1, and the other countries will be ranked proportionally in between the countries receiving the highest and the lowest score.

These seven indicators – objectives, target groups, finances, provision models, benefit generosity, activation conditions, and sanctions – thus form the input of the benchmark model. As explained in chapter 2, these will be presented and aggregated using a radar-chart and the Surface Measure of Performance (SMOP). Table 4.1 on the previous page summarises the main component of the choice of the indicators as discussed in this sub-section.

4.2. Process indicators

Including indicators measuring the different elements of the process of the programs to be evaluated is one of the most novel aspects of this research. As mentioned before, this will allow for better understanding of why certain results are being achieved, beyond only relying on the input or the output of the program. Similarly to the input, two levels of the process can be defined. The first way the process can influence the results of a program is through effective and efficient institutional settings for the delivery of the services. The assumption here is that results will be positively affected if the manner in which the organisations deliver the program is fostering an environment in which participants can easily receive the most appropriate assistance. The second level of the process which can positively influence results is the effect of specific interventions on the participants themselves. Within this micro-level, it is expected that certain interventions which have been shown to be linked with effectiveness will result in a larger number of participants returning to the labour market. And here again, a multi-disciplinary approach will be used for the selection of the relevant indicators and the measurement of their performance, with the legislation in each country also playing a large role in the measurement of the process of each programs.

The institutional-level process is measured by looking at the extent to which the governance model for the delivery of employment services is expected to be efficient and effective. The first element which will be
analysed is the extent to which each program fits the four types of provision models for employment services as discussed by Mosley and Sol (2005). These are: the bureaucratic model, the Management by Objective model, the preferred-supplier model and the quasi-market model. First, the bureaucratic model is one in which the government is itself taking care of the delivery of the program. Actually, contracting-out is only used in niche services and competition is not part of the incentive mechanisms. This is different in the second model, the Management by Objective model, since a high level of competition is found between the different public or private agencies delivering the services. In this model, contracts are used to regulate the relationship between the central government and the delivering public agencies. The last two models are both based on market mechanisms. In the quasi-market model, the state is purchasing the services through a tendering process in which price or quality will be the main criteria on which providers will compete. On the other hand, the preferred-supplier model defines countries in which the state will procure from providers whose quality has been proven in the market. As can be seen, both models rely on private providers for the delivery of employment services and use contracts as main steering instruments. The main difference is found in the degree of competition created by outsourcing, which is lower in the preferred-supplier model and highest in the quasi-market model.

However, no direct assumption can be made on the type of partnership which should lead to better result, as there is very little evidence available on this (see for example Hasluck and Green (2007, p. 32) on the UK). Indeed, as also pointed out by Mosley and Sol (2005), even though many countries have reformed their provision model towards market-based approach in the idea that this would increase their efficiency, the success of such reforms is often debated. Indeed, as also pointed out by many researchers on the subject (see for example Considine, 2005; Bredgaard and Larsen, 2005; Struyven and Steurs, 2005; Van Berkel and Van der Aa, 2005), privatisation is also not automatically guaranteeing effectiveness and efficiency. Actually, some countries such as the Netherlands have even seen a reversal of the privatisation of service delivery (there in the case of employment services for social assistance). However, what can be assumed to positively affect the results of the program is the extent to which the delivery of the services is tied to the achievement of some performance level. It can be expected that the more the payments for the service rendered will be linked to the performance of the program, the better the results will be in terms of rate of return to the labour market. Obviously, it is most likely that the analysis of this factor will quickly follow the typology described by Mosley and Sol (2005), where the more the provision model is market-based, the more there are concrete
performance agreements between each actor. However, this does not need to be the case, and consequently the indicator will avoid making such an assumption. Hence, in the benchmark the country which presents a form of delivery which is most accountable to the result of the program will receive a score of 5, and the country with the least performance incentives built in its delivery model will receive a 1. The other countries will be ranked with respect to this range created in between these values.

The other six indicators in this sub-section will now have to do with the process of the program and its impact on the individuals taking part in it. Similarly to the assumptions concerning the micro-effect of the input, micro-economic studies can help find out which elements of the process can contribute to high rates of return to the labour market. Many evaluations have been performed on different groups of projects and on different indicators which are part of this larger benchmark exercise. Typically, these micro-economic studies in fact only focus on one aspect of the program, for example as will be discussed next, the timing of the start of the activities. Putting these different studies side by side thus gives an indication of which elements lead to the success of work-based employment programs.

The first micro-level process indicator to be discussed is the **timing of the start** of the program. Ochel (2005), who performed one of the rare reviews of the determinants of success of Work First programs, indeed concluded that the timing of the start of the program is an important contributor to the success of these programs. From a more general point of view, Lechner and Wiehler (2007) have shown that in the case of the Austrian benefit scheme, the effectiveness of labour market programs deteriorate the later they start within the unemployment spell. The authors note that their result corresponds to the findings of two other international studies from Sweden (Sianesi, 2004) and Germany (Fitzenberger and Speckesser, 2005). These findings coincide with the discussion in the previous chapter on the intervention strategy of the programs, where it was also said that individual must also be prevented from loosing their ability, willingness and access to work. This move from the first quadrant of the model to any of the other quadrants can most quickly be prevented by a rapid start of the work activities. This way, the skills and knowledge as well as the incentive structure of the unemployed can stay as close as possible to the state in which they were previous to the benefit claim. From the perspective of a later start of the program, it was also discussed when looking at the indicator of the target groups that the long-term unemployed have a smaller chance to find a job because of increased barriers and skills depreciation. This thus also
means those programs which start only after the unemployed has been claiming benefits for a long time are also expected to have lower rates of return to the labour market. Also, an early start of the program was associated with a smaller lock-in effect in a Danish study by Graversen (in Ochel, 2005, p. 85). The lock-in effect will be discussed in more details within the next process-indicator (the length of the program), but it is important to note here that it means that participants in a program will delay their return to the labour market for after the program has ended. By starting earlier, it can be hypothesised that the perceived need to finish the program will be smaller because the person is only unemployed for a short time. That way, there may be less chance that the person waits for the end of the program to look for a job. Consequently, the score of 5 will thus be given to the program which starts at the moment the claim is made, and the score of 1 will be given to the country where the program starts the latest after unemployment benefits have been claimed.

One problem with such a measurement is that the start of a claim and the length of a claim are not automatically linked with how long someone has been inactive on the labour market. While one person might have just lost a job and claiming a benefit the next day, another person might have actually been in an education program or caring for children, which means that this person’s work experience is less recent (if existent at all) than the other person. This problem is usually not present when looking at unemployment insurance schemes, because it can easily be assumed that claimants have been working prior to making their claim. However, as will be seen in part 2, most work-based employment programs actually operate within the social assistance scheme, where claimants do not necessarily have recent work experience. Clearly, if data is available on this aspect it should be taken into consideration in the benchmark. However, we are not so much concerned here with the timing with respect to previous work experience as with timing with respect to the start of the claim. Early intervention is expected here to have better results than later intervention, no matter the actual length of inactivity of the participants. Variations in the composition of the participants with respect to the work history prior to their claim are then assumed to balance out differences within program with regards to that aspect.

The next process indicator measures and compares the length of the programs in the benchmark. As explained by Van Ours (2004), locking-in effects can arise when participants take part in an employment program and therefore reduce job search intensity. This reduction can arise from different reasons, such as a lack of time and resources to search for jobs and difficulties in attending interviews while participating in the program. Besides these practical explanations, one can also imagine that
some participants would rather complete a program before proceeding to intensive job search. In this respect, the more effective the program is expected to be, the more the participant will have the incentive to complete it, especially if the program will allow the participant to find better jobs with better salaries. This could surely be the case of training programs, and indeed many researchers have found that training programs have a significant locking-in effect (see amongst others Martin (2000) on the inefficiency of training programs). In fact, programs which only bring their benefit if they are fully completed (such as training, where not completing will imply not receiving the certificate) should be designed with much care in order to minimise locking-in effects.

It is clear that locking-in effect can be diminished by making sure the participants in the program both have the resources and the incentives to pursue intensive job search during their participation. This was thus measured earlier with regard to activation conditions within the program (as part of the input indicators), and the actual amount and type of job search assistance received by the participants will also be measured as part of the output indicators. Nevertheless, the length of the program has also been associated with locking-in effects, since the longer the program will be, the longer the effect of lower job search intensity will be felt. Furthermore, a short program can make sure that participants remain focused on finding a job, and not see the program itself as their primary focus for that period time. Indeed, Van Ours (2004) and Lechner and Wiehler (2007) concluded that the shorter the program will be, the smaller the locking-in effect will also be. Hence, the shortest program duration will receive a score of 5 in the benchmark, while the longest a score of 1. Clearly, there is a lower limit in which short duration implies better results, since at the extreme a program lasting one day would not be expected to be more effective than a program lasting one month. This is because programs must allow some time for the participants to learn and gain experience and this will need at least some weeks before being fully realised. However, the precise turning point at which the effect of short duration will not be observed anymore is unknown and not discussed as such by researchers looking at this topic. This turning point can nevertheless be assumed fairly realistically to be under the boundary of 3 months which is the shortest program duration within the benchmark, as will be shown in part 2.

The third indicator of the process on a micro-level is the number of hours spent in the program each week. Here, the assumption is that the more intensive the program is in terms of hours per week, the better the results will be. Intense contact with the participants will indeed make sure that the incentive structures set up by the program achieve the desired goals.
As explained in the previous chapter, work-based employment programs intend to make the choice of work more interesting than not working, and this could be done by either increasing skills or preferences, or both. Access to work through networks and job search assistance was also mentioned as the third important factor. The number of hours can influence these three aspects in various ways. First, the intensive intervention will have a positive impact on the quality of the skills acquired through training and work activities. Obviously, the more intensive the program will be, the more experienced the participant will become and the greater the learning effect will be from training. Also, the closer the contact between the participant and the decision-maker regarding the type of skills to be learned, the easier it will be to move the participant from an unable state to an able state. Similarly, Ochel (2005) suggested that intensively monitoring participants on a personal basis is said to have a positive influence on the results of work-based employment programs. Furthermore, work-based employment programs are also expected to alter preferences towards not working in the way that some disutility is experienced compared to not participating in the program. As discussed earlier when looking at the activation conditions, this disutility will mean that participants will increase their job search in order to be able to leave the program (OECD, 2000. Ch.4). It can be assumed that this disutility will increase as the number of hours per week in the program increases. The signalling effect will also be affected by the number of hours per week, as potential employers can interpret this as a proof that the unemployed is able to work for a substantial amount of hours per week. The longer the hours, the stronger the program can be seen as a realistic work-test. Lastly, designing the program such that it requires many hours of participation per week will have a larger effect on fraud, since it will restrain the time available for undeclared activities (Bodganor, 2004). This all points towards a ranking in which the program with the longest hours per week will be expected to have the best results, and thus receive a score of 5. The program with the least hours will thus receive a score of 1, with the other countries ranked in between.

The fourth indicator refers to the type of work-activities to be performed as defined by their working environment in which they take place. In fact, Ochel (2005) has also found this element in many evaluations to be of great impact on the success of work-based employment programs. The working environment indicator refers to whether the work-activities are part of the regular labour market, in which the demands of the tasks match those of the regular labour market, or whether they are part of
what can be defined as an “intermediate labour market”\(^4\). The distinction “real work” versus “simulated work” can maybe shed some light into this distinction. This means that activities either take place in the private for-profit sector or public sector in which the work-activities are part of the regular activities of the employer, or in the public or non-profit sector in which the work-activities are part of activities for which there is no market-demand for. This distinction is very clear when comparing work-activities which take place in a private business to work-activities taking place in community organisation solely based on voluntary work. However, when it comes to work within the public sector, this distinction is however more difficult to make. Work-activities can in some case be fully incorporated in the regular activities of the organisation, clearly resembling the situation found within private employers. It can also be that the work-activities are not part of the regular activities of the public organisations, and are purely additional and created especially for the purpose of the program.

The filling-in of this indicator will thus require more interpretation than simply looking at the sector in which the work-activities will take place. Nevertheless, the programs in this benchmark all show characteristics which will make it possible to rank them on a scale on the extent to which the work-activities are connected to the regular labour market. The closer the activities take place to the labour market, the better the results are expected to be. As Ochel (2005) also mentioned, when the activities are in a private work environment, the participants face the same conditions as they would in a real job. His study showed that a “real” work-environment which was similar to the demands of a job on the regular labour market influenced the results of the projects significantly more than when the activities where taking place in a simulated work-environment, which was ran publicly. Additionally, a real work environment should facilitate the contact with a network of potential employers. According to network theory, informal weak ties can be very productive (Granovetter, 1974). Ochel showed that the skills learned in a private work environment where much more valued by future employers than the skills learned in a public work environment. The future employer perceives the participant as more productive and is therefore more quickly willing to hire him or her. Furthermore, Ochel also mentions that the disincentive effect of private work environment is larger, since the

---

\(^4\) For a definition of Intermediate Labor Market, see Finn and Simmonds (2003), Intermediate labour market in Britain and an international review of transitional employment programs, Department of Work and Pension research report #173; as well as Marshall and Macfarlane (2000), The Intermediate Labor Market, a tool for tackling long-term unemployment, Joseph Rowntree Foundation.
participants are more likely to consider the working conditions harsher than they would be in a public program. This thus makes them more prone to accept a real job when the work-based employment program takes place in a private work environment. As a result, by offering a work-environment which is closest to the regular labour market, the work-based employment program is able to intervene on each dimensions of its three-dimensional intervention strategy, by increasing the ability to work, the willingness to work, and the access to work. The country with work-activities the closest to the regular labour market will thus receive a score of 5, while the country with work-activities the most removed from the conditions found on the labour market will receive a 1.

Another micro-level process indicator is the type and amount of rewards given to the participants in return for their participation in the program as a whole or for the work-activities in specific. As discussed in the section on the input-indicators, the participants are actually claiming an unemployment benefit when their participation in the program is required (as already mentioned it can be so that this claim lead to an immediate start of the program, or that the program will start a while after the benefit is being received). Three things can happen when the program starts: first, the participants can continue to claim the regular benefit at the same level as all other claimants while being in the program, second, a bonus or complement can be added to the regular benefit as a reward for performing work-activities, or third, the participant can actually be removed from the caseload of the benefit scheme to be employed in the program and thus receive a regular (subsidised) salary. It is obvious that the amount of the rewards which is additional to the regular benefit amount should make a difference on the willingness of the participants to take part in the project. Also, giving a salary in return of the work can be much more rewarding for the participants and take much of the stigma attached to the social assistance claimant status. This positive signal can also increase the access to the labour market for the participants.

The influence of benefit levels on the results of the programs were already taken into account as part of the input indicators. It was argued that lower benefit levels would increase the chance that participants exit the program to return to the labour market, mainly because of the financial incentives arising from the difference between the benefit level and the prospective wage received on the labour market. However, it was also mentioned that some arguments point to the fact that barriers to work can be created by low benefit levels. Indeed, this view is centred on the idea that too low benefit level will distort the attention and energy of the unemployed away from job search to other necessary activity such as
finding low-cost shelter and food, or repaying debts. Also, poverty and unemployment itself can be a cause of many socio-psychological problems (such as depression) which lower the chance of finding work. It is thus clear that giving generous rewards for participating in the work-based employment program could alleviate some of these negative effects stemming from low benefit levels. In fact, the combination of low benefit levels with high rewards for work-based employment programs would create a situation which was contemplated as being optimal by Robert Solow. His opinion was that there should be equilibrium between self-reliance and altruism, where work is rewarded and unemployment not being punished (in Beaudry, 2002, p.23). Hence, the country with the highest rewards for participating in the program will thus receive a 5, and the country with the lowest will receive a 1. Since we are interested here in the extent to which additional incentives are given for participating in the work-based employment program, this indicator will be measured in terms of percentage increase with respect to the benefit level alone. Also, the specific type of rewards will be taken into consideration in the ranking, where regular salaries are preferred over benefits for their signalling effects as well as their effect on the motivation and self-respect of the participants. However, it is expected that higher rewards will be found in programs where salaries are given, such that rankings will be preserved.

The sanctioning mechanism in place within the benefit scheme legislation was also part of the input indicators. Harsher sanctions were expected to have a positive influence on the results since this will act strongly on the willingness to work of the participants. However, it can be the case that this legislation has left room for the implementing body to choose from a broad range of sanctions which they can apply. Indeed, as noted by the OECD (2000, p. 138), only measuring the strictness of the sanctioning mechanisms does not provide complete information on the impact these sanctions can have on unemployment levels because information on how these sanctions are implemented is also necessary for such analysis. For example, the OECD shows that while sanctions for refusal to accept a job are not so severe in Denmark, they are much more often applied than in Belgium, where sanctions are very harsh and thus reluctantly used (OECD, 2000, p. 134). Clearly, the sanctioning procedure can have an important impact on not only the sheer number and type of sanctions applied, but on the actual threat effect these sanctions pose to the claimants.

First, it can be assumed that the greater the flexibility in the procedure, the better the sanction will be able to reach its desired effect and be fine-tuned to the severity of the breach committed. Indeed, it was mentioned
by Vonk (2008, p.15) that the main condition for sanctioning procedure in
work-based employment programs is that these are fair and are
proportional. Hence, the first sub-indicators regarding the sanctioning
procedure will thus be the amount of discretion available to the delivering
body of the program in term of applying sanctions or not, and in varying
the length and duration of the sanctions. However, flexibility should not
equal complete surrender to the arbitrary decision of the case manager in
the sanction to be applied. This would then lead to an unpredictable
system which could quickly be deemed as unfair because highly
dependent on the professionalism of the case-manager. Therefore, a
second sub-indicator will be added, that of the degree of formality of the
decision making process. A high degree of formalism is expected to have a
positive influence on the result of the program since this foster a
trustworthy procedure in which rules are systematically applied, leading
to expectable results (i.e. rule of law). Moreover, a high degree of
formality can also be associated with easy and clear appeal procedures,
which also contribute to the fairness of the procedure.

All in all, the combination of both flexibility and formality would lead to a
sanctioning procedure which would be deemed fair, both by claimants
and by those applying the sanctions. This in fact not only applies to
sanctions with regard to social security benefit, but to sanctions and
punishment in the law in general. Indeed, three principles are
fundamental to criminal justice, and thus also to administrative law and
its sanctions (Brouwer et. al., 2004, p. 449). The first one is the legality
principle, where a punishable behaviour can only be so if this behaviour
is prohibited by law. This principle is clearly related to the indicator of
formality as described above. The second one is the culpability principle,
where behaviour can only be punished if the person has intentionally
committed the punishable behaviour. This also related to a formal
sanctioning procedure as mentioned above. The third principle of
criminal law is the requital principle, where the reason for punishment as
well as the degree of punishment must find reason in the objective gravity
of the accomplished offence. This is similar to the indicator of flexibility
of sanctioning procedure as included in this benchmark.

The concepts of formality and flexibility are certainly related to each
other. A formal system where clear rules are present for the application
of sanctions can easily be expected to also lead to a flexible system in
which different sanction levels are given for different degrees of severity
of the breach. However, this does not need to always be the case. Indeed,
it could be so that a very formal system only contains one level of
sanction, and this would lead to a highly predictable system, where one
could easily know that the consequence of a breach would be to receive a
sanction, although this sanction could not be adjusted to the severity of the act. Hence, while it can be expected that formality and flexibility will go hand in hand, this might not be the case for all countries, and different relative levels of formality and flexibility can be expected. Hence, both indicators will be given a rank of 1 to 5 and the country combining the highest degree of formality with the highest degree of flexibility will thus receive a score of 5, and the other countries will be ranked accordingly, with the lowest level of fairness being awarded a score of 1.

Summarizing, a total of seven indicators will measure the extent to which the process is likely to be optimal for good results. Table 4.2 summarises the choices made in this benchmark of the process of work-based employment programs.

Table 4.2  Performance indicators for the process of work-based employment programs

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Sub-indicators</th>
<th>Score = 5</th>
<th>Score = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode of Delivery</td>
<td>Mode of provision</td>
<td>Strongest focus on efficient delivery</td>
<td>Weakest focus on efficient delivery</td>
</tr>
<tr>
<td></td>
<td>Performance pay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timing of start</td>
<td>Number of weeks after claim before program starts</td>
<td>Quickest start</td>
<td>Latest start</td>
</tr>
<tr>
<td>Length of program</td>
<td>Number of months</td>
<td>Shortest program</td>
<td>Longest program</td>
</tr>
<tr>
<td>Hours per week</td>
<td>Number of hours per week in program (total for all activities)</td>
<td>Most hours</td>
<td>Least hours</td>
</tr>
<tr>
<td>Work environment</td>
<td>Private for-profit</td>
<td>Closest to regular labour market</td>
<td>Farthest to regular labour market</td>
</tr>
<tr>
<td></td>
<td>Public/regular</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Public/simulated</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Voluntary sector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rewards</td>
<td>Benefit only</td>
<td>Highest reward as % of regular benefit, and salary as a reward</td>
<td>Lowest reward as % of benefit (only benefit as reward)</td>
</tr>
<tr>
<td></td>
<td>Bonus on benefit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Salary</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Amount of reward as % of benefit level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sanctioning procedure</td>
<td>Formality</td>
<td>Most formal and flexible</td>
<td>Least formal and flexible</td>
</tr>
<tr>
<td></td>
<td>Flexibility</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.3. Output indicators

As explained in chapter 2, the output refers to what is being “produced” by the program. These are the direct effects of the program on the participants, such as for example increases in skills level, changes in preference for work, or a better access to job opportunities on the labour market. Actually, it is through this output that the impact is realized, that is to say, entering the labour market. Since objectives can also be formulated in terms of output, it is important for the calculation of the effectiveness and the efficiency to include these in the benchmark. Consequently, the three-dimensional intervention strategy defined in the previous chapter will also provide a framework for those output indicators. Through this three-dimensional intervention strategy it was described how programs attempt to increase ability to work, willingness to work and access to work. The output indicators will thus measure the extent to which those products of the programs have been achieved.

The first output indicator will look at the effect the program has on increasing skills levels. This will thus imply looking at the training activities which are part of the work-based employment program. Training activities can be either formal or informal, as part of more traditional class-room methods or else happening on-the-job. It was also discussed in the previous chapter that the work-activities themselves are also providing for on-the-job skills training. Nevertheless, this training component which is inherent of the work-activities of the work-based employment program can also be complemented by a formal skills training program. In general, an argument against “workfare” is that these programs foster an under-investment in training, since these are more expensive measures for which the return-on-investment is believed to be less than other provisions such as work-activities and monitored job-search (Beaudry, 2002). However, this benchmark looks at programs where training can be an important part of the provisions, and this hypothesis of under-investment in skills will thus also be discussed through this indicator.

Thus, this indicator will give a measure of the increase in ability to work through the formal skills training provisions which are part of the program, since it is already a given fact that in all program work-activities take place in which some informal training occurs. The country with the most extensive training program both in terms of hours per week as well as formality and transferability of the skills learned will receive a score of 5. The other countries will be ranked based on the relative position in terms of training with this country. The country with no formal training component will receive a score of 1. In the later case, this does not mean
that the ability to work is not at all being improved, since skills-improvement can also be done “on the job” through the work activities. Nevertheless, a low score does mean that, relative to the other country in the benchmark, less is being achieved in terms of skills training, since all other countries also provide some type of “on the job” skills training through work-activities.

It should be noted however that formal training activities will also affect the other dimensions of the intervention strategy. As it was discussed in the previous chapter, increasing skills will also have an effect on the willingness to work since a higher wage will be able to be earned because of improved productivity levels. Also, since employers use education levels and other qualifications as a signal for the level of hard and soft skills, increasing skills through formal training will also increase the access to the labour market. Thus, while this indicator is a closer proxy for increases in ability to work, it also has strong links with the willingness to work and access to work as well.

The second output indicator relates more directly to the goal of increasing access to work. This will be measure by the type and amount of job search assistance. One of the determinants of success for work-based employment programs described by Ochel (2005) is the appropriate mix of workfare activities and job-search assistance in the program. The author concluded that the most effective programs are the ones that focus the most attention of job search assistance as opposed to focusing on work-activities only. As also mentioned by Lightman, Mitchel and Herd (2005), an important group of welfare claimants has significant barriers to work which will require thorough interventions beyond what can be offered while taking part in workfare activities. The amount and quality of job search assistance will thus be the indicator of the increase in access to the labour market, although other provisions can also increase access to work. These other provisions are amongst others, as discussed through the process indicators, providing work experience on the regular labour market and paying out a regular salary instead of a benefit. This increases the access to the labour market since they give the participant the chance to have a recent work experience on the résumé which is valued by potential employers (see Ochel, 2005). Nevertheless, the provision of job search assistance throughout the duration of the program is a direct way to influence the access to work, while those other provision act indirectly through changing the prejudice potential employers may have against unemployment benefit claimants.

This Job search assistance indicator will be made of two components. The first component is the intensity of the supervision and guidance received
by the participants. This supervision and guidance can have two different sources, that is to say, the (private) provider of the program and the government body which is implementing the program (and contracting-out the direct service delivery). Indeed, it may be so that, even though a private provider is responsible for the day-to-day supervision by its own case-managers, the government organisation implementing the program still also provides some case-management to the claimants. The later is especially true when issues of eligibility and sanctioning are raised. The second component of this indicator is the number of hours per week which is required or made available through the program. For example, the program can ask the participants to spend one day per week on searching for a job. The country for which both of these components is the highest, thus were supervision and guidance is the most intense and where job search requirement are included in the weekly planning in the program will thus score a 5. Countries has very little supervision and guidance for the participants and no time allocated to job search in the program will on the other hand score a 1. All other countries will be ranked according to their relative position to these two situations.

The third aspect of the three-dimensional intervention strategy was the increase in the willingness to work of the participants. It was discussed in the previous chapter how unemployed individuals will be unwilling to work in a situation where it is more “satisfying” in terms of utility to choose for being unemployed and receive a benefit. By eliminating the option of enjoying all free time and not working, the work-activities were argued to have an effect on the willingness to work. However, the activation conditions and accompanying sanctioning mechanisms were also said to increase the willingness to work because of the financial consequences inactivity will bring. Sanctions have indeed been shown to have a positive effect on the rate of return to the labour market by authors such as Abbring, et al. (2005), Van den Berg, et. al. (2004), Svarer (2007) and many more (see previous section). In terms of output, the “products” of the program, the increase in the willingness to work will be difficult to measure through a proxy. While the number of hours in the work-activities can be seen as having a certain disutility effect which will increase the willingness to work, this is not an assumption which will hold for all participants in the program. Also, looking at the amount of sanctions handed out will only show that a proportion of the participants was first unwilling to accept a job or participate in the program, but will not say anything about the effect of the sanctions on those who did not receive one. It has in fact been shown in some micro-economics studies that sanctions can have an important ex-ante effect, where participants will comply more with the program because of the existence of the sanctions (see section 4.1 on sanctions). However, all these effects have
been taken into account in the benchmark as part of the input and the process indicators. It was then assumed that sanctions and work-activities will influence the rates of return to work of the program through increasing willingness to work. Seeing the lack of micro-evidence which would be able to show increase in willingness to work, the actual link between the input/process and the impact will not be calculated as part of the output indicators.

Nevertheless, the quantity of sanctions which are yearly used in each program should still be measured as part of the output indicators. This would not be in order to calculate a proxy for the increased willingness to work, but rather a product itself of the increased conditionality these types of programs bring about. Indeed, it is very much the question whether the strengthening in activation conditions and the sanctions that foster compliance to these conditions is actually being implemented at the street-level by case-managers. This would mean that the increase in conditions and sanction would only take place in the legislative field, and thus could be seen as a way to make symbolic law. Symbolic law can be defined as a type of law in which the strongly normative message in which political values are conveyed is the core of the law, and this law not meant for the actual enforcement of its conditions (Aubert, 1969). By looking at the extent to which the combination of the sanctioning procedure and the sanctioning mechanisms result in sanctions actually being used, the hypothesis of symbolic law can be tested. Nevertheless, in the case where sanctions are not part of symbolic law by that they are truly being used by case-managers, it is still difficult to gauge what the number of sanctions given out actually says. For example, it is difficult to tell whether a small amount of sanctions has to do with the individual participants in the program and their propensity to breach, or whether is has more to do with the sanctioning procedure and the institutional setting in which sanctions take place. To answer this, sophisticated micro-economic experiments would need to take place in which the characteristics of the participants and the characteristics of the sanctioning procedures would be taken into account. What can be expected is that the actual number of sanctions given out per year in a program is both a combination of its participants’ propensity to breach and its own decision-making procedures with regard to the sanctions. Also, some important factors can come into play, such as the presence of financial incentives from giving out a sanction to participants. Consequently, a qualitative discussion on the number of sanction used in relation with the process indicator of the sanctioning procedure should therefore be very informative on the extent to which sanctions are truly being used or whether they are only part of some change in discourse without change in use of instruments.
The last indicator as part of the output benchmark is the coverage of the program. Because the objectives of employment programs have to do with either increasing ability, willingness and access to work, or fostering the return to the labour market, a measure of how many unemployed are being assisted through the program is also necessary. This is because the actual implicit or explicit macro-level objectives of the program will be to decrease the total number of unemployed in the country. A program where the measure of the output and the impact would be very positive should also be evaluated against its quantitative effect on the number of claimants who benefit from these provisions. This way, if the objective of a program is to raise the skills level of the unemployed, the combination of the output-indicator concerning the training activities and this indicator of coverage will give a good measure of the effectiveness of the program in terms of raising skills of the unemployed. This is also true when looking at impacts, which as will be explained next, is measured in terms of percentage of the participants who returned to the labour market, and not in terms of the total caseload.

Adding the indicator of coverage to the analysis of the impact will give a good idea of the real impact on total unemployment. Moreover, this indicator will also be very informative of the sheer size of the programs which are being compared, and this quantitative aspect is a very much needed addition to the qualitative evaluation of these programs. This indicator will be measured in terms of total yearly caseload within the relevant benefit, and not in terms of target groups, in order to reflect the impact a choice of a narrow target group will have on the coverage rate of the program. With respect to ranking, it is understood that the larger the program will be in terms of the total unemployment caseload, the larger the effect of the program on unemployment levels will be. Hence, the country with the largest coverage will receive a score of 5, while the country with the smallest coverage will receive a score of 1.

Table 4.3 on the next page gives an overview of the performance indicators which will be used to measure the output of the work-based employment programs.
Table 4.3  Performance indicators for the output of work-based employment programs

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Sub-indicators</th>
<th>Score = 5</th>
<th>Score = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>- Hours per week</td>
<td>Most hours and most formal/transferable skills</td>
<td>Least hours and least transferable skills</td>
</tr>
<tr>
<td></td>
<td>- Formality and transferability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job search assistance</td>
<td>- Intensity supervision</td>
<td>Most intensive supervision and most time allocated for job search</td>
<td>Least intensive supervision and no time for job search</td>
</tr>
<tr>
<td></td>
<td>- Hours per week job search in program</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of sanctions</td>
<td>- Percentage of participants sanctioned per year</td>
<td>Qualitative analysis in combination with other indicators</td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td>- Number of participants sanctioned per year in % of total yearly caseload</td>
<td>Highest coverage rate</td>
<td>Lowest coverage rate</td>
</tr>
</tbody>
</table>

4.4. Impact indicators

The difficulties encountered when evaluating labour market programs where introduced in chapter three, where it was shown that even though much progress has been made regarding evaluation methodologies, still not one method stands out as providing the best estimate of the performance of a program. There is indeed a large body of literature relating to the proper way to calculate the impact of a program (see Smith, 2000; Kluve, 2006; Martin, 2000; and De Koning, 2007; Thomas, 2008, for reviews of different methodologies and evidences).

As explained by amongst others, Smith (2000) and Thomas (2008), the main problem faced when attempting to evaluate active labour market programs is that one can only observe an individual once, either as this person takes part in a program, or that this person does not. The difficulty lies then in estimating what had happened in the other non-observed stage. Thus, while gross impact only look at whether a person entered or not the labour market after participating in a program, net impact is concerned with calculating only those participants who did enter the labour market through the program, but would not have done so otherwise. As a result, various statistical methods have been developed by econometrists to estimate this net impact (see Smith (2000) for a review of the most common methods). More concretely, what needs to be eliminated are the impacts which would have occurred anyway.
even in the absence of the program (deadweight loss) or impacts which
do not create additional employment in the economy (substitution and
displacement effects) (see chapter 2, section 2.2.1 for more on the dead-
weight lost, substitution effect and displacement effect; and also Calmfors,
1994). For example, if a program only attracts highly motivated and only
recently unemployed individuals, one can expect this program to have a
high rate of return to the labour market. Consequently, proponents of
net-impact studies argue for the removal of this advantage given to the
program by first estimating the number of persons who would have
found a job on their own, and then subtracting this number from the gross
rate of return to the labour market. These types of evaluations are
therefore essential for the proper evaluation of labour market programs if
one is interested to know the impact it has on total unemployment. It is
also very useful for comparing the impact of different programs since
variation in target group is taken into account in the statistical model.
The benchmark should therefore rely on evaluations of the net impact of
the programs in order to compare their rates of return to the labour
market.

However, the methodologies to calculate this number of participants who
would have found a job on their own are still being developed and are
varying from one evaluation to the next (see De Koning, 2007, p. 5). This
thus means that first, many evaluations of labour market program do not
calculate net impacts as these require expert knowledge on micro-
economic evaluations models. The data requirement is also much more
complex when attempting to calculate net-impacts. This thus means that
these net-impact calculations are actually rather seldom and surely not
systematically performed for all labour market programs in many
different countries. This should surely pose some problem for the data
collection in part two of this benchmark. Indeed, only the UK and
Australia performed net-impact estimations of their work-based
employment programs5. Second, when looking at evaluations from
different country, it is very much likely that different models will be used,
which will be based on different assumptions and different variables.
This would lead to estimates of net impact being not fully comparable,
because of the differences in methodologies. This is also the case for the
studies from the United Kingdom and Australia (see Beale, Bloss and
Thomas, 2008 and DEWR, 2006, p.8). Without going to much into details
into the debates around the appropriate methodology, it still is
noteworthy to mention how several researchers are starting to discover
important flaws in commonly used statistical models. For example,

5 Canada, the Netherlands and Switzerland only present data on gross impact,
and even those are in the case of these countries not easily accessible.
Lemieux and Milligan (2006) argue that the difference-in-difference models which are often used for net-impact evaluations are very sensitive to the choice in the control group such that they may perform poorly when this group is inadequately chosen. Another example is Lalive, Van Ours and Zweimüller (2008) who compared two dynamic estimators (matching estimators and timing-of-event estimators) and found that they resulted in different net-impact calculations.

Because of this major lack of evidence and this incompatibility in methodologies, only gross impact will be used in this benchmark. Nevertheless, the use of gross impact is less problematic when the evaluation is based on an international benchmark based on the entire policy-chain. First, we are here not concerned with the added value of program within the range of instruments used within a country, but we are concerned with the relative success of programs compared to other similar programs around the world. Second, differences in the target group and other individual pre-selection effects will be taken into account by the input-indicators of the benchmark. These will thus contribute to the explanation of why better results are being reached. In addition, external factors will take into account such effects as high economic growth or low employment standards which could also explain why a certain program is reaching higher return-to-work ratios. In conclusion, it is clear that while net impacts would be much more desirable to include in this benchmark, the sheer lack of available studies in some country, as well as the lack of a common methodology makes this first-best solution unavailable. Nevertheless, since the benchmark model is able to take into account some of the differences in the participants as well as differences in external factors, these gross impacts will be analysed in the light of these indicators. This will bring the analysis of the impact closer to the net-impacts than would otherwise be possible with a benchmark which would only take into account impact indicators. The need to remedy this lack of data availability will be discussed later on in the next chapter where conclusions on benchmarking labour market programs will be made.

The first impact indicator will thus be the rate of outflow to work, as measured by the number of participants who find a job either during the program or at the end of the program, as a percentage of the total number of participants who started the program. One important feature of this indicator is that it only looks at outflow to unsubsidized jobs, and not to outflow out of the benefit, as the objectives of the programs are not solely to diminish caseloads, but to do so through the labour market. A participant leaving the program to withdraw from the labour market would not be considered as having reached the objectives of those
employment programs. Also, as will be shown in the second part of this research, it is often so that participants in work-based employment programs leave the benefit scheme to become employees of the program. This would mean that those participants automatically exit the benefit, but this should not count as a positive outcome for the programs. Concerning the ranking of the countries against the scale of 1 to 5, it is obvious that the country with the highest rate of outflow to work will receive a 5, and the country with the lowest rate will receive a 1, with the other countries ranked proportionally in between.

Nevertheless, an important limitation to consider in the evaluation of the impact is also the sustainability of this return to the labour market. There is undeniably a risk that work-based employment programs stimulate the unemployed to enter precarious and low-paid jobs (Beaudry, 2002). The danger of this is the creation of a revolving-door phenomenon, where cycles of employment are alternated with cycles of social assistance (see Handler (2007)). While in the short-term the objective of Work First seems to have been reached, a long-term approach would show little improvement in the economic self-sufficiency of the individual. An evaluation of the Wisconsin Works program in the United States showed that 52% of the participants who left the program from 1999 through 2002 participated again in the program by mid-2004 (Department of Workforce Development, 2005). The fact that most participants in the Wisconsin Works program entered mostly low paid jobs is reflected by the data gathered on income levels of the participants after taking part in the projects. In 2004, 20% (with tax credits included 33%) of the former participants had an income above the US poverty-line (DWD, 2005). This meant that a majority of the participants did not manage to sustain economic self-sufficiency in those 3-5 years. The average income reported for tax purposes of the former participants was 9,000 US$ (DWD, 2005). Clearly, the majority of the participants in the Wisconsin Works program are likely to become “working-poor”, that is to say, individuals who hold a job but still do not manage to have an income above the poverty line.

These results are specific for the Wisconsin Works program but can surely be expected in other similar Work First programs. As Peck and Theodore (2000) pointed out, there is a general link between Work First and deregulated flexible labour markets. Additionally, they fervently argued for the fact that Work First created a ‘forced’ labour supply for contingent jobs, where the unemployed has no choice to make a transition to a low-waged job (Peck and Theodore (2000)). Accordingly, besides measuring the status of participants at the end of the program, a measure of the sustainability of their return to work should be taken into account. However, this type of data is also much more seldom in program
evaluations. When available, a measure of the status of the participants 3 to 6 months after leaving the benefit for a job should be taken into account in the analysis of the program, even though comparisons between all countries in the benchmark might not be possible due to lack of data for some countries.

Even though most programs have the principal objective to foster the return to the labour market, there is surely an implicit macro-level objective to decrease the total caseload number, and thus lower the unemployment rate. Total exit from the benefit, not only to take-up a job, should thus also be included in the measure of the impact of the program. Since some programs automatically create the exit of all participants in their program, another way in which this indicator can be calculated is to look at the number of participants who return to the benefit after having participated in the program. A measure of the rate of return to the benefit will thus be included in the benchmark of the impacts of the programs. Since this indicator is measuring failure instead of success, a score of 5 will be given to the country with the smallest rate of return to the benefit, while the country with the highest rate will score a 1.

The next indicator of impact also measures the effect on the total caseload, but this time the preventive effect of the program will be measured through the rate of inflow prevention. Besides fostering exit to work, work-based employment programs have a considerable threat-effect on claimants. This threat-effect is especially due to the disutility experienced from work-activities which are required from the participants. However, as said earlier when discussing the output with respect to the willingness to work, this disutility effect will not mean that all potential participants will rather leave the benefit than having to participate in the work-based employment program. As also mentioned by the OECD (2000, ch.4), some participants will still choose to take part in the program but will significantly increase job search efforts in order to leave as soon as possible. Some characteristics of the programs increase the threat-effect and these were discussed throughout the input and process indicators. For example harsh sanctions or not receiving any additional benefit for participating in work-activities can have a higher threat-effect for the potential participants (see Ochel, 2005). This threat-effect thus means that the total caseload is not only reduced from the fact that participants find a job on the labour market, but that some claimants decide to leave the benefit before the program will be made mandatory for them. This can be either from the start of the claim, or after a while, as discussed in the section on the process-indicators. A measure of the proportional number of unemployed exiting the benefit just before the work-activities must start will therefore be included in the benchmark of the impact of the
program. Since this prevention effect with help to attain the macro-level (implicit or explicit) objective of reducing the unemployment rate, a large prevention effect will receive a large score on the benchmark. In fact, as the other indicators, the country with the highest rate of prevention will receive a 5 and the country with the lowest rate will receive a 1, with the other countries ranked proportionally in between.

The measure of the prevention effect of the programs raises an important issue when looking at work-based employment programs. In fact, one major limitation of work-based employment programs emerges from the fact that the choices presented to the individuals are wider than only either taking part in the employment program or working on the labour market. The economic choice model presented in chapter 3 showed that work-based employment programs being mandatory, the individuals were then faced with the choice of working or taking part in the work-based employment program.

In reality, other choices are present for the individual. For example, the unemployed can have other sources of income which will make the welfare benefit unnecessary. These income sources can range from informal-market activities or even illegal activities, to family members or friends willing to financially support the unemployed. In addition, work-based employment programs being ran by the municipality, the unemployed can move to a municipality where the welfare requirements do not include the participation in Work First. Another option for the unemployed is also to simply accept to live under the poverty-level income. The presence of these other options imply that the incentives structure set up to influence the willingness to work also have an impact on the willingness of any of these other options.

Nevertheless, it is clear that the objective of the programs is not simply the reduction of the social-assistance caseload but its active participation in the labour market. When the unemployed opt for any of these alternative options, this objective is thus not met. Critics claim that especially young people just disappear from the official statistics. They are left with no job or benefit and society wrongfully accepts that part of the young literally fall through the bottom of the system (see FNV, PZC (2006)). This fact is of concern since it should not be forgotten that the primary goal of Social Assistance is not the activation of the unemployed but clearly the alleviation of poverty. By creating work-based employment programs which give the incentive to the unemployed to accept a level of income below the poverty-line, these programs thus undermine the effectiveness of Social Assistance. An important indicator to measure would thus be the level of income of all those exiting the
benefit system either as a result of the threat-effect of the programs. However, this type of data is hardly being collected by the public employment services, since these claimants are not part of their administrative data anymore, and would need extensive tracking through surveys or through the income tax system. Nevertheless, in the event this type of data is available, this indicator should surely be added to the benchmark. Through this indicator, it would be possible to know whether this threat-effect has indeed resulted in a large number of unemployed finding a job on their own, which would also diminish the deadweight loss of the programs and thus also raise the effectiveness of the program.

Table 4.4 Performance indicators for the impact of work-based employment programs

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Score = 5</th>
<th>Score = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of return to work</td>
<td>Largest rate of return to work</td>
<td>Smallest rate of return to work</td>
</tr>
<tr>
<td>Sustainability of job placement</td>
<td>Largest number of participants still in job after 3 months / 6 months</td>
<td>Smallest number of participants still in job after 3 months / 6 months</td>
</tr>
<tr>
<td>Rate of return to benefit</td>
<td>Smallest rate of return to benefit</td>
<td>Smallest rate of return to benefit</td>
</tr>
<tr>
<td>Rate of inflow prevention</td>
<td>Largest rate of inflow prevention</td>
<td>Smallest rate of inflow prevention</td>
</tr>
</tbody>
</table>

Table 4.4 shows these four indicators of the impact as discussed so far. These will allow measuring the effectiveness of the work-based employment programs with regards to the objectives of fostering the return to the labour market of the unemployed as well as decreasing the unemployment caseload in general. It will also take into consideration the negative effects the programs can have on the sustainability of this return to the labour market, by looking at effects such as the revolving-door phenomenon and the increase in working-poor through these programs.

4.5. External factors

The next group of indicators is of a rather different nature of the other indicators. These refer to contextual elements which are found outside of the domain of social security but still have an impact on the design, the implementation, and on the results of work-based employment programs. As explained earlier when looking at the impact, it is important to take
these into account especially since gross impacts of programs found in different countries are being compared. It could thus be so that the very good results of a country would not have much to do with the input or the process of the programs, but rather to those contextual factors, such as a growing economy. Even though most evaluations only look at the influence of external factors on impacts, they can nevertheless also have an influence on all the other elements of the policy chain. For example, economic growth can also influence the objectives of programs since assumptions about how easy it should be to find a job will be different in that context. The ranking of these indicators will therefore have a different interpretation than the other rankings as described earlier. Since the effect of the external factors will be found on all the policy-chain, the ranking will not be based on the hypothesis of the influence the indicator will have on the impacts. This is because while an external factor indicator could have a positive influence on the impacts, it could be so that it has a negative influence on the rest of the policy chain, which would then not be visible from the ranking. Instead, rankings will be only based on the value of the indicators, such that 1 will be for low levels of the indicator and 5 for high levels. It will then be easier to make separate hypothesis for the influence of these ranks on each part of the policy-chain separately. It should thus be clear that the analysis of the impact of the external factors will not be of quantitative nature, but will rather take the form of a qualitative analysis. It is thus not so that this benchmark will statistically correct the impact indicators for diverging levels of external factors. That is not to say that such an exercise is not useful, on the contrary. In fact, Mosley and Muller (2007) have recently benchmarked 141 public employment service offices in Germany and did correct the performance level of each office for exogenous factors such as economic condition and client characteristics. This resulted in a benchmark in which ranks truly represented the performance level of each office, since external effects were removed from performance levels. It is very unfortunate that such a method cannot be used for this benchmark. The very small number of units in the benchmark indeed does not allow for such techniques to be applied to the benchmark in this research. Nevertheless, by proceeding to a qualitative analysis of the external factors which could explain important differences in design, implementation and results, this research does recognise the importance of taking these factors into account. This is especially true when conclusions and recommendations are made on which programs should serve as a best-practice example to other programs.
Even though the strength of the economy provides for some good examples of what external factors can be, these are surely not the only type of factors which are meant here. Indeed, a broad array of institutions, rules, norms, market mechanisms will have an influence on the policy chain. The question here is thus how far to go in taking these into account. In fact, although it might be an interesting exercise to list all possible factors which will forge the policy-chain of work-based employment programs, the feasibility of the benchmark must be kept in mind. Thus not only the measurability of these factors must be taken into account, but mostly the availability of the data which are internationally comparable is a must. As with each category of indicators, the balance between the efforts required to gather the data and the added-value of the information for the understanding of the program must be maintained. It should thus be kept in mind that the benchmark is not meant as a mean to measure all possible aspects of the program, but to highlight the most important determinants of success and failure of these programs. The external factors have the objective to bring broader perspective into the analysis of the programs and allow a more thorough understanding of the different factors affecting its policy-chain.

The logical way to achieve this is to first define three broad categories of external factors which can summarize and encompass a broader range of factors. The first and most obvious category is the economic context, since this has without any doubt a major impact on the number of people unemployed in the country and the ease at which they will be able to find a new job. The second category is the juridical context, since many laws outside the social security laws will have an influence on some aspect of the design and implementation of the programs. Laws also shape the labour market in which the participants in the program will enter. At last, the political context in which the program takes place will also be taken into account, since this will have an important influence on the design of the programs and the way it is being implemented. The specific indicators for each of these three categories will now in turn be discussed.

The first category of external effects is the economic context of the country or region in which the program is taking place. First, the economic context will determine how many people are unemployed and who they are (mostly young people, mostly older worker from declining industries, low skilled workers in general, etcetera). This will of course then influence the target group of the program. As already discussed when looking at target groups (section 4.2), the unemployment rate of the young unemployed is much more volatile than the general unemployment rate, which means that in economic downturn they will make up a larger part of the stock of unemployed but that when the
economy regains momentum, they will also find jobs in greater numbers. This will of course have an impact on the input of the program as defined earlier, and surely affect the results of the program. But also the process may be influenced, since it might be easier to find employers in the private sector, or even in the public sector, that are willing to take part in a work-based employment program if they have more difficulties filling their job openings. On the other side, it might be so that in times of economic down-turn, employers might be more interested in having employees who receive a wage subsidy. The precise effect of the economic context on the process is thus difficult to predict. The largest and most obvious effect of the economy on the program will be on the impact it will have in terms of return to the labour market. It has indeed often been stated that reforms introducing work-based employment program often take place in periods of favourable economic conditions, which would make the program appear more successful than it really was (Beaudry, 2002).

Hence, a measure of the extent to which the country was going through a period of economic growth will be added to the external factors, which will allow to correct for this positive effect on the results when comparing the different countries. It will also allow to highlight large differences in economic context which could have an impact on the input or the process indicators. First, this indicator will be made up of the GDP growth and the change in unemployment rate with respect to the previous year. Those two indicators will give an indication of the strength of the labour market at the time the participants where looking for jobs. If the GDP growth was low and the unemployment rate increasing, one can expect that it would be much more difficult to find a job than in the case where the GDP was growing fast and the unemployment rate declining. To these relative measures of the economy will be added an absolute indicator of the general labour market conditions, that is to say the unemployment rate itself. Indeed, all other things being equal, a higher unemployment rate would indicate a labour market where there would be less demand for labour, especially labour supply coming from the unemployment caseload. The country where the combination of these three indicators will represent the best economic condition will thus receive a rank of 5, while the opposite will be true for the country which will score a 1.

The second type of external factor has to do with the juridical context in which the program takes place, outside of the social security legislation which is taken into account in the input indicators. As with each other indicators, it has been acknowledged that work-based employment programs, just as a public policy, is formed by the law. This is true when
we look at both the instrumental functions of law as well as its normative functions. This meant that the social security legislation which was in direct link with the programs, mostly through the legislation on the unemployment benefit, were the source of information for many of the input and process indicators. However, other laws outside of this social security field can have an indirect effect on the programs, shaping its input, process, output and impact. The range of laws which affect programs is in fact very broad, ranging from contract-law to laws regulating the age at which children should start school and until when they should remain in school, to fiscal regulations. Listing and measuring all these laws and regulations would however be impossible in any benchmark, and here one must look at the juridical indicators which are most likely to have the most influence on the programs.

With respect to the impact the program will have on the return to work of its participants, the labour law context will surely have a significant effect. Since work-based program either directly operate on the regular labour market or operate very close to it, employment regulations will have a great impact on the working conditions of the program participants both during the program and after leaving to take-up employment. In fact, in the cases where the work-based employment program provides a regular job on the labour market, these employment protection legislations have already been taken into account when discussing the sanctions in the program since this will be the channel which will be available for sanctioning participants. In those case the employment protection legislation is first having a direct effect on the participants through the sanction mechanisms, and a second indirect effect by influencing the labour market which their enter after the program has stopped (defined by the moment the subsidy is not given anymore to the employer).

Authors such as Peck and Theodore (2004) argue how the type of jobs which are taken-up by the participants in work-based employment programs are low-paid and short-term contracts, leading to very precarious situations. Quoting Peck (2001, p.6) in its introduction to his famous book “Workfare States”: “Stripped down to its labour-regulatory essence, workfare is not about creating jobs for people that don’t have them; it’s about creating workers for jobs that nobody wants”. It should however be mentioned that this analysis concerns mostly programs found in the United States, where the level of employment protection is very low, and minimum salaries also very low. But seeing the fact that this can also very much apply to any other developed country, an indication of the insecurity of the situation of all those who return to the labour market will be included in the benchmark. The OECD has calculated an often used
index of the strictness of the employment protection legislation in its member-countries (see OECD, 2004, chapter 4). This index is based on three components: 1) protection of regular worker against dismissal (core component), 2) specific requirements for collective dismissal, and 3) regulation of temporary forms of unemployment. One important drawback for this index which the OECD has put forward is that many indicators within this index can be arranged differently through the collective labour agreements. To this measure provided by the OECD will be added one last component of the labour market situation, namely the minimum wage which would be valid for those who would have been part of the program (in some country this wage is lower according to age, or if one is a new employee).

The question here is how these factors should be ranked in the benchmark. In fact, the precise effect of EPL on the labour market is ambiguous (OECD, 2004, ch.4). On the one side, it is expected that a low level of employment protection will mean that it will be easier to hire new employees which are coming from the unemployment benefit. That is because employers take into account the eventual firing costs when hiring new employees, and low firing costs will thus be less of a hindrance (OECD, 2004, p. 76). On the other side, the sustainability of this return to the labour market is expected to be much lower in case EPL is not very strict because of these low firing costs. Also, a low minimum wage is expected to foster new hire, but at the same time this will mean that the recently employed person might not have much financial security. This ambiguousness is exacerbated by the fact that these hypotheses are not fully empirically proven. In fact, as the OECD discusses, data availability and comparability are major obstacles for cross-country longitudinal studies of the effect of EPL on the labour market. As a result of these ambiguous effects, the ranking in the benchmark will not be given in terms of effect on the impact of the program. This clearly highlights one drawback of this method, in which ranking needs to be based on a single common criterion, or otherwise this ranking will not be comparable from one indicator to the next. As a result, the ranks will be from low to high (instead of 1 to 5) and only a qualitative analysis will be performed. The programs will the be analysed in the light of the possible effect low EPL can have on higher short-term exit rates but lower levels of sustainability of the exit to work.

The employment protection legislation and minimum wage indicator are the two external factors which are selected to represent the juridical domain in this section of the benchmark. That is not to say that these are the only laws and rules which have an effect on the programs. On the contrary, the number of juridical arrangements which can be linked to
work-based employment programs seems more to be unlimited. One can think here of indicators of the costs of hiring new employees, the transaction costs involved in contracting private providers for service delivery, the rules on occupational health and safety within the programs, and working time regulations, to name a few. However, it is also important to consider the data availability and comparability of such indicators. Furthermore, it can be questioned whether there would be enough variation from country to country to be able to make some inference about influences of these rules on the employment programs. This small variation makes it then unnecessary to include in the benchmark, since these could not explain large differences in input, process, output or impacts. As an example, it is surely the case that the international legislation on the prohibition of forced labour does play a role in forging work based employment programs. However, the direct causality link between the interpretation of these rules in each country and the design of the programs would be difficult to make. This would be first because this type of information is not easily available, and second because the variation from country to country will not be so pronounced as to highlight differences in a significant manner. Nevertheless, when benchmarking labour market programs, one should pay careful attention to which legislation may present a significant degree of variation from country to country such that including this indicator would benefit the evaluation of the programs. In this case here, employment protection legislation and minimum wage legislations were deemed to have potential explanatory power.

The last indicator of external factors will be used more to explain the choices in design and modes of implementation than to explain the results reached by the program. This indicator will relate to the political context of the country, which is expected to have an important influence on the social security system of the country in general, but also on the timing of the introduction of the work-based employment program and initial choices made within this introduction phase. In fact, a discussion of the objective of the programs will have to include some notions of the political context in order to fully understand why certain choices are made above others. Political scientists around the world are researching the effect political parties and differences in ideology can have on the welfare state, which are complementing studies that mostly concentrate on welfare state regime typologies. When Esping-Andersen (1990) wrote his famous typology of welfare states regimes, the coincidence of political parties and welfare regimes made it easier to leave the analysis of the political context outside of consideration. Indeed, the liberal welfare state of the UK was lead by Margaret Thatcher, corporatist Germany ruled by Kohl and Sweden was ruled by the social-democrats. However, this clear
cut distinction did not hold over the last 20 years and looking at the political context surrounding work-based employment program will help understand design and implementation more than if only welfare state regimes were taken into account. The first indicator of the political context will therefore be the party which introduced the program in the first place. The second indicator will denominate if and when this party changed and the shift in ideology which this change brought about. These two external factors indicators will thus allow a comparison of the different work-based employment programs while understanding the political context in which the development of the program took place. However, links between political context and the results of the program are not possible to be made, since it is unclear how the ideological background would affect rates of return to work. It could be inferred that if a majority of the countries citizen votes for social-democratic parties that this could indicate a corporate culture in which social responsibility could be important, thus implying that employers would be keener to employ unemployment benefit claimants. However, this inference between voting behaviour and corporate social responsibility is not very strong. On the other side, one can expect that the political context does have a more direct influence on the input and the process of the programs. It will be assumed here that liberal and conservative governments would be more inclined to set up programs which would be more focused on sanctions and conditionality than would social democratic governments. The process within the program would also be expected to be containing more negative incentives, such as long working hours and little extra financial benefit from participation. As was discussed earlier in the section concerning the input and the process, some of these negative incentives are expected to have a positive effect on the results (sanctions), but others are not (low extra rewards for participation). In this way, no clear assumption can be made about the influence of the political context on impacts, but most indicators of the input and process will be better explained by looking at this external factor. The ranking will thus be qualitative.

Table 4.5 on the next page shows the selected three external factor indicators which influence the whole of the policy-chain, but are found outside of the social security domain in which the programs take place.
Table 4.5  Indicators for the external factors of work-based employment programs

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Sub-indicator</th>
<th>Score = 5</th>
<th>Score = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic context</td>
<td>- GDP growth</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Changes in unemployment rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Unemployment rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour law context</td>
<td>- Strictness of EPL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Minimum wage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political context</td>
<td>- Party introducing the program</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Changes of party</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.6. Conclusions

This chapter illustrated one of the main difficulties of the benchmark methodology, namely that all indicators need to be ranked in one direction with respect to their influence on the results of the programs. However, such inferences might not always be straightforward to make. A considerable amount of consensus must be found in the literature before one can make an assumption on the way in which a certain indicator will influence the results. While it was the case that enough evidence could support the choice for ranking with most indicators, a few indicators where much more difficult to rank. This was surely the case for the following two indicators, the benefit level as well as the sanction level and duration. For both these indicators, a U-shaped relationship with the impacts seemed to be the most realistic assumption that could be made. Indeed, very low benefits and very harsh sanctions could have the reverse effect they attempt to achieve by actually hindering the unemployed in entering the labour market. However, the precise turning point of this U-shaped relationship remains not know empirically. Consequently, the ranking will be in the first instance based on the assumption of a linear relationship between the indicators and the exit-to-work ratios. Nevertheless, the analysis of these indicators will be done with extra care, and vigilant attention will be given to the possibility of having made the wrong assumption by implying a linear relationship between the two indicators and the impact. In addition, with respect to the indicator representing the benefit level, an indicator of rewards given through the program will allow to account for the assumption that financial incentives within the program itself can also have a positive effect on the rates of return to work. As a result, the combination of the assumptions for the benefit level and for the rewards within the program is expected to provide for a realistic assumption on financial incentives.
Furthermore, the interpretation of the indicator on the quantity of sanctions was also difficult to place with respect to the ranking of the indicators. As it was discussed, a high number of sanctions could be seen as an indicator that many participants were putting a higher effort in finding a job, since requirements to search for a job were then respected. On the other side, a low number of sanctions would indicate that most participants are actually already following these requirements, such that sanctions are not needed. The inclusion of this indicator was nevertheless deemed very useful, not in relation to the impacts of the program, but to be able to make a complete evaluation of the sanctioning mechanism of the programs. Indeed, sanction will now be measured as an input, as a process and an output, which will be very interesting to analyse.

At last, it was also clear that the relation between the policy-chain and the external factors is made-up of a number of different positive and negative influences on many parts of the chain at the same time, such that the total effect on the program is very difficult to foresee. Also, due to the small amount of cases in the benchmark, statistical correction of the performance of the programs for differences in external factors is not possible. It is nevertheless acknowledged that the research would have greatly benefited from such a possibility. Other benchmarks in which the number of cases is large enough for such statistical analysis should surely attempt to use these types of corrections. Here, the analysis will have to remain of qualitative nature, by putting the results of the benchmark under the light of the economic, legal and political context these programs take place. The value of this qualitative analysis lies in the fact that conclusions and recommendation will be more cautious as to the direct application of lessons from one program to programs found in other countries. Also, since external factors can also vary through time, the analysis will also give an indication of how performance levels can be expected to vary as the context within which these programs takes place will change.

Clearly, social benchmarking relies heavily on the available research on the effectiveness of single indicators for the performance of labour market programs. Benchmarking thus serves as a way to use the information from these micro-level evaluations in a way which complements these studies, and surely not to replace them. Appropriate research on the impact of different elements of the input and process is thus very much necessary, and these are even stimulated thought social benchmarking. Much attention in future research should be given to strengthen the hypothesis needed for social benchmarking such that the ranking of performance indicators is based on stronger assumptions. Especially research on the effect of very low benefit levels as well as very harsh
sanctions is needed in order to better understand whether there is a limit to the extent to which these indicators can be expected to improve chances to find a job. In the context of the increased use of activation conditions in order to make active labour market even more “activating”, such information on the limits of sanctions and negative incentives is clearly needed.