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.
12. Conclusions and recommendations

The research question of whether benchmarking is a useful instrument for evaluating labour market programs is certainly being answered positively. In part 1, a social benchmark model was developed which can be adapted to evaluate a wide range of active labour market policies. This benchmark model provided the theoretical answer to the research question of how benchmarking could be useful for evaluations. It was showed that by using this new model, benchmarking is able to combine qualitative and quantitative analysis in such a way that enough details are provided to those willing to understand why a certain level of results are being reached. This social benchmark model thus fills an important gap by providing a methodology which is able to analyse the performance of labour market programs in an informative way.

The second part of this research has shown that much could be learned from an extensive discussion and comparison of the input, process, output, impact, and external factors of work-based employment programs. This exercise provided a good example of how benchmarking can assist in making better programs which are able to increase the effectiveness and efficiency of labour market policies. This second part had in fact served to verify the theoretical answer to the research question, following the idea that “the proof of the pudding is in the eating”. From this, some remarks can be made about some shortcomings of the model which were discovered while benchmarking work-based employment programs.

The first shortcoming has to do with the weights of the rankings within each element of the policy-chain. In this model, all indicators had equal weights, meaning that it was assumed that their influence on the impact would be also equal. This however is most probably not the case. The difficulty in choosing those weights lies in the unavailability of reliable hypotheses on the strength of the relationship between each indicator and the impact with respect to each other. For example, if one would be able to assert that the influence of sanctions is twice as high as the influence of the benefit level with respect to outflow to work, one would be able to give a weight to the indicator for sanctions which would be twice as large as that of benefit level. The absence of clear hypotheses on the relative importance of each indicator in the input, process and output benchmarks meant that weights were not used, even though one could justify their use from a more intuitive point of view. This benchmark model could thus be improved by using weights on its different indicators, and further research is necessary in order to choose those weights.
The second limitation of the benchmark as performed in part 2 of the research has to do with the fact that strong causal relationships between each indicators could not be pinpointed with certitude, that is to say, using statistical analysis. Although this could be solved by adding many more programs to the benchmark in order to increase the number of observations, the benchmark model itself might constraint such a choice. Indeed, the choice of benchmarking the whole policy-chain, instead of only just results, means that a relatively large number of indicators needs to be collected and analysed. While this does not formally prevent a large number of cases being included in the benchmark, the model does put an important constraint on the amount of time and resources needed to be invested in data collection and analysis. This model is thus by design more made for a stronger qualitative analysis than a quantitative one. This does mean that most statistical tests will most likely not provide significant answers in this model.

At last, the lack of comparable data on an international level should be also mentioned, although this issue is also faced by other evaluation methods. More specifically, the model can difficulty cope with situations where data is missing in one or two of the cases, since this missing data will in one way or the other influence total rankings. This benchmark model could therefore be improved by finding a way in which missing data can best be compensated for. Also, when the data that was available was not directly comparable, different estimations and calculations where made in order to compare those data. This has however lead to many conclusions having to be drawn with caution, due to the fact that the data used was an approximation. As data collection in the different countries is being improved, and the international comparativeness of this data fostered, this benchmark model will be able to strengthen its analysis.

All in all, the social benchmark model developed in this research was successful in identifying the source of performance gaps within the different work-based employment programs it benchmarked. Improvements in terms of weights, available data, and statistical analysis would contribute to improve the lessons to be learned from good and bad practices, but these limitations do not undermine the fact that benchmarking remains a useful instrument for the evaluation of labour market programs.

Furthermore, one important point to be acknowledged in this concluding chapter concerns the actual use of benchmarking results by policy makers. Indeed, although research on active labour market policies does have a place of its own in academic debates, the question whether these
evaluations do end up assisting policy-making should be asked. By knowing how evaluation results can become more useful for evidence-based policy making, social benchmarking can be better tuned to the needs of policy-makers. In fact, much research has already been done in the United States on the use and on the dissemination of evaluation results, and those findings can be useful to keep in mind when building a social benchmark model. Firstly, Greensberg et al. (2000) studied the dissemination and utilization of three major innovations in social policy in the United States in the 1990's. The authors mention how since 1981, and similarly to the situation created by the devolution of social programs in many European countries, the 50 states of the country can be seen as a laboratory for experiments on social services delivery. They claimed that in such a setting, if an experiment is positive, it should disseminate to other states and be implemented. The researchers looked at the impact of three experiments, and saw that rather than having concrete or instrumental effects, evaluations had more a conceptual or enlightenment effect. The effect was thus not dramatic on policies, but played a role in deliberation in most states. The same findings were made by Sol and Castonguay (forthcoming, 2009) on the dissemination and use of the results of a benchmark of 49 work-based employment programs in Dutch municipalities. Greenberg et al. (2000) also showed that other than evaluations, sources of policy change were: politics, pure evolution of policies where change was the next logic step, change in economic situation, as well as bandwagon effect of popular programs (however not based on evaluation). In sum, knowledge that a policy is logical and politically appealing is all that is required, not empirics. Evaluations then reinforce thoughts and directions and thus add to the knowledge base. In fact, for most evaluations, the information on the process of the program, on how the program was operating in the field, was much more important than the empirical evidence of the effectiveness of programs (Greenberg et al. 2000).

Hence, it should be clear that social benchmarking should not only provide reliable “numbers” on the net-impact of programs. Social benchmarks of labour market programs should focus on addressing the issues that are currently important to the policy-makers and law-makers. Such issues have often to do with different ways to manoeuvre within the given parameters of the (national) programs, making indicators of input and process crucial to the usefulness of the benchmark. In more concrete terms, with respect to work-based employment programs, there is a clear need for more information on the consequences of increased conditionality within the benefit scheme and the programs. Little is known about the impact of low benefits, harsh sanctions, stringent work and extensive job search requirements. Many hypotheses can be made on the
limits to which activation can go before it starts to actually decrease the effectiveness of the programs. This social benchmark has shown that there is evidence that these hypotheses are true and they can explain why some of the harshest programs in the benchmark had much lower performance levels. It should thus be able to answer some of the crucial questions that many policy-makers and law-makers are asking themselves concerning mandatory work-based employment programs.

Another issue which is of high importance is the change in what is being regarded as a positive result for active labour market policies. Indeed, effectiveness and efficiency are no longer solely measured with respect to rates of return to the labour market, but also with respect to the impact on the lives of the participants further than only their labour market status at the end of the program. There is thus a shift in the focus from short-term exit rates to long-term measures of sustainability. There is also a shift from the idea that “any job is better than no job” towards looking for jobs which are able to provide a stepping-stone for the unemployed. Moreover, this also implies that the objectives of active labour market policies are increasingly considering the socio-cultural aspects of the return to the labour market, where social inclusion and social participation have an explicit role to play in the programs. The financial and social situation of those leaving the benefits, either because they found work or any other reason, is thus increasingly of interest to both researchers and policy-makers. Social benchmark should address these new ways of looking at impact by incorporating such alternative measures of impacts in their performance indicators. By doing so, social benchmarks will be useful for policy-makers and law-makers whose concerns are quickly broadening outside of what is traditionally looked at in micro and micro evaluations.

At last, benchmarking is also a very useful evaluation tool in the context of never-ending welfare reforms, where the issues at stake and the design parameters of programs are constantly changing. Considering this, the crucial next step for performance management is to adapt benchmarking methods to the new reality of rapidly changing contexts. This can easily be done by creating dynamic longitudinal benchmarks, which can rapidly adapt to new circumstances. With information technologies making huge leaps in what is possible in terms of data collection and dissemination, such an adaptation is realistic. Benchmarks could then contain data which would be collected at frequent periods in time, and also make it possible to change the variables measured according to new needs. Dynamic longitudinal benchmarking would then properly fulfill the needs for evidence-based policy-making in times where adaptation to new socio-economic contexts is key to efficiency and effectiveness.
Evaluations can also be seen as the link which turns the policy-chain into a full circle. Knowing the level of impacts allows to question whether the objectives of the program where realistic and appropriate. The evaluation of the effectiveness of programs means that new objectives might be given to the program. Then, by knowing which elements of the input, the process and the output lead to which types of impacts, the whole policy-chain might be adapted to better match new objectives. This cycle of setting up objectives, inputs, processes, creating output and measuring impacts is thus directly in relation with what the evaluation model is able to achieve.

The objectives of active labour market policies, the data collected on how the intervention strategy is being implemented, and the indicators which are being used as a success or failure criteria are all linked with each other. Improving evaluation models can thus have an important influence on the design of programs, since new impacts can be highlighted which can then become a more formal part of the objectives in the program. Being part of the objective of the program will also mean that data will be collected on this element, allowing a better evaluation of their impact. Objectives, monitoring systems, and evaluation models thus clearly need to be developed conjointly in order to benefit from each other. Seeing the increased questioning of the broader impact of active labour market programs not only on employment but on other aspects of the lives of individuals, a broader perspective in both the programs being developed and the evaluation models being used is desirable. This broader perspective on activation should take into account a larger array of impacts than what is traditionally being evaluated in micro-level or macro-level studies. However, without proper monitoring systems, even an evaluation model based on a broader perspective will not be able to draw the conclusions is intends to make. As policy-makers and lawmakers find many of their questions concerning the overall success of activating labour market programs unanswered, data collection on a larger range of indicators will be stimulated. As a result, social benchmark models like the one developed in this benchmark will then be able to explore the relationship between the design and results in a more profound manner.

In order to guide these developments in the use of social benchmarks, this research makes the following five recommendations:

- Social benchmarks should take into account the whole policy-chain, allowing analysing not only which program performs the best, but also making it possible to understand why such a performance level is being attained.
The content of the social benchmark should look at movements from unemployment to employment from a multi-disciplinary angle. A broad perspective needs to be given to what is being considered a successful transition to work, as this should not be to the expense of overall welfare.

The needs of policy-makers and law-makers in terms of guiding their choices when designing employment programs should be met through this benchmark, since benchmarking should lead to programs being improved. Determinants of success and failure need to be identified such that programs can learn from each other.

Further research is necessary in order to better understand the theory being benchmarking as a tool for learning from other practices. Theories from the field of organisational learning need to be better adapted to public policy, and in particular the field of labour market policies where legitimacy is increasingly being questioned. How benchmarks should be used within an organisation to help decision-making at all levels, from the top-manager to the case-managers, needs to be further looked at.

On a more practical side, technical solutions with respect to weighting ranks, ranking missing data, using incompatible data sources, and finding statistical evidence within a benchmark with a small number of observations need to be explored. This would allow the benchmark to make stronger inferences on the causal relationships between its indicators.

Since the social benchmark model developed in this research was also tested on actual programs, some recommendations can be made with respect to the designing of work-based employment programs. These recommendations are drawn from the previous chapter, which brought all the elements together from each benchmark of the policy-chain elements.

Focusing on negative incentives within the intervention-strategy will work backwards and reduce the effectiveness of work-based employment programs. Investing in a proper level of positive incentives is necessary, and adequate training, job search assistance, a private work-environment, decent income in the program, and a short program duration are important success factors.
The entire policy-chain needs to be included in the intervention strategy of work-based employment programs. A high level of initial conditions, but combined with poor choices in the process, will not result in a successful program. Similarly, the right choices in process-elements will not be sufficient in creating effectiveness if the initial conditions within the benefit are detrimental to its success.

Proper data collection on the outflow-to-work is essential in order to evaluate those programs, and was lacking in many programs. Furthermore, other impacts such as job sustainability and threat-effect of the programs are impossible to evaluate due to the lack of available data. The same is true for the number of sanctions used, which is in many countries poorly monitored. Without proper data collection on all these important elements of work-based employment programs, it is impossible to know what is the broader effect of those programs on the lives of the unemployed. As these types of data will be made available, it is important that further research be undertaken to verify if the determinants of success as discussed within this benchmark still hold within this broader perspective on activation.