Evaluation of the Berlin Innovation and Technology Support (Germany)

1. Introduction

The evaluation study applies to the evaluation of a programme of support in Berlin. The programme consists of five measures and has been a cornerstone of the innovation policy in the city-region in question.

In addition to evaluating the five measures / programmes in question, the evaluation also assesses the linkages and interplay between the measures. The evaluation features a combination of quantitative (analysis of administrative data; an online survey of beneficiaries) and qualitative (in-depth interviews; case studies) data collection and analysis methods. In assessing the impact of the measures, the evaluation distinguishes between (1) an analysis of the achievement of immediate results in terms of the funding provided based on a micro-economic analysis, and (2) a longer term, macro-economic analysis. It provides results on the level of individual measures and the programme overall.

The evaluation was useful on various levels. It was of major importance for the programme management, as it gave justification of money spent. The evaluation led to a better mutual understanding of achievements and challenges in other programmes. For the main programme manager there were no big surprises in terms of effects, but the overall neutral proof and learning experience was extremely valuable and allowed for the application of any adjustments needed. This also led to learning across different sub-programmes and measures in a way that was not possible in the day to day operation of the programme or from the monitoring activities. The evaluation and its process achieved a shared understanding and joint perspective as to the relative role of the programmes and their contributions to the overall policy goals. This also allowed for a re-structuring of programmes (namely, the fusion of two programmes), which had been considered beforehand, but the data and evidence gathered in the evaluation gave the justificatory basis for actually implementing this change.

At the political level the evaluation was welcome as a means to communicate achievements to the outside world (via a formal press conference) and to the Regional State Parliament. Further, the evaluation proved that the tax effects of increased economic activity stemming from the programme exceeded the inherent programme costs, which added to the legitimacy of the programme.

The evaluation deliberately did not look at efficiency and programme management, and the programme owner made very clear that he wanted to avoid the evaluation being transformed into a strategic consultancy exercise with advice on future directions and foci rather than an evaluation of the programmes' impact.

2. Description of the evaluated measure

The objectives of the evaluation were very broad; the programme is a cornerstone of the region's innovation policy. The measure is an umbrella programme with five measures, each having specific sub-goals. Overall, the umbrella programme seeks to improve and capitalise on the innovation potential of firms located within the region. The overarching goal is to better utilise the existing research capacity in the region for the region's economy, thus the instruments in the measure are very much orientated towards transfer and cooperation activities. The long term objectives are to achieve more employment and growth through improving the innovation capabilities of firms and science, the improved competitiveness of firms and more efficient structures in selected competence fields.

The five programmes and their main objectives were as follows:

- Support of research, innovation and technology: to improve networking around innovation and increase technology intensity and innovation capacity in the economy as basis for more competitiveness.
- Technology Coaching Centre: to achieve the build up of long term structures through advice and training, management support more generally and development of soft skills.
- Future fund: to achieve the build up of sustainable cooperation structure between science and industry with a view towards the development of new products and lighthouse projects – e.g. projects that stick out and have signalling effect in terms of ambition and feasibility – in selected competence areas.
- VC funds: two funds to secure innovation financing and improve the set up of firms.
- Innovation Assistant: to support the build up of long-term employment of scientists in firms and to speed up of knowledge transfer.

The target groups for these measures are highly overlapping: the first two measures focus exclusively on firms – often with a focus on SMEs - while the last three support academic institutions as well.

3. Designing the evaluation study

The State Law compels the administration to do an evaluation, but does not prescribe its form. The previous evaluation of the programme was done 5 years ago, thus it was timely to conduct another one to check for developments. In addition, the State Parliament wanted to see the justification for the budget spent. In parallel to the requirement to design the evaluation, there is also a continuous monitoring process in place in order to back up the large evaluations (based on regular surveys), and individual questions are tackled over the years with short term analyses.

The design of the evaluation was done mainly internally; the programme management wanted a multi-level analysis (at the micro-, meso- and macro-scales). The comprehensive meso-economic part (focusing on competence areas) could not be conducted in accordance with the expectations of the programme manager (see below). The programme management wanted to avoid the possibility that the evaluation would turn into a strategic consultancy rather than resulting in hard evidence of the programmes' effects. This occurred with the previous evaluation which had been conducted by another prominent consultancy group.

Further, a circle in the ministry and a few external people – trusted evaluation experts – were involved, and at some stage an external expert gave some input (prior to tender) to get a feeling of what could be reasonably expected in terms of the scope and depth of the evaluation in respect of the budget available to the programme owners.

Despite various consultations, the programme owner and manager himself formulated all the evaluation questions. The programme management further made sure that the money for the evaluation was found outside the programme budget, as they were anxious to maintain a clear distinction between budgets for the target group (programme beneficiaries) and budgets for evaluations.

On the basis of a socio-economic situation analysis the evaluation's main objective was to analyse the impact of the programme and to provide a comparison to other regional programmes in Germany. It was intended that it should examine the strategic fit and appropriateness of the individual measures and their interplay to support the structural policy goals of the region. A further major objective was to test the

sustainability of the programme and to formulate recommendations for future developments of the innovation support system.

The concrete questions formulated in the Terms of Reference can be summarised as follows. The evaluation was tasked to analyse and discuss:

- The acceptance of and demand for the measure, structure of the support payments (in terms of recipients);
- how and to what degree the measures support the structural goals of the regions and how are budgets allocated to competence areas;
- the interplay of measures within the programme and external to it (including national and EU measures);
- the way in which intermediary institutions manage the measures or take them into consideration in their own activities;
- a potential deadweight loss;
- a potential gap in the support system that could be covered by further measures;
- potential modifications to the schemes, especially as regards interplay of measures;
- comparison to other regional programmes and drawing of lessons;
- the suitability of the programme to achieve the goals of the master-plans;
- a potential strengthening of gender orientation.

The Terms of Reference hinted at methods that should be used for the evaluation, but remained open as to their concrete application. They mentioned a number of indicators as a tentative and not conclusive list:

- number of projects supported
- number of projects terminated
- support quota (share of eligible companies supported), degree of coverage of target group
- number of advice sessions
- number of collaborative projects (Verbundvorhaben).

Note: In the macro-and microeconomic analysis further indicators were used by the evaluating team, those are mentioned in the next section.

In addition, the Terms of Reference demand a multi-level and time sensitive evaluation: *Short term goal achievement* and attribution of changes to the measure should be assessed through tender document analysis (which allows analysis of the types of failures addressed and thus suitability of the funding instruments) and written surveys and interviews, taking into consideration multiplier effects through private or other public co-investments.

The issue of data need and availability was discussed in depth at the early stages and even during the preparatory stages of the evaluation. This then led to some changes for the meso-level analysis (see below) and adaptation to the survey design; in one instance existing survey material was to be used rather than conducting a new survey.

4. Implementing evaluation: methodology and process

The evaluation was conducted by external evaluators, selected through a tender procedure. One important principle for the programme manager was to try to remain within the European tender threshold, which somewhat limited the size of the evaluation. There are two reasons for this: First, the international tender process was regarded to be too cumbersome. Second, from previous and current evaluation experience, the programme manager had learned that national and local consultants are better equipped to understand the context of the federal state in question.

The programme management followed the national commissioning law which did not in any way compromise what they wanted to do in terms of content and selection procedure, etc. They applied an open procedure and – as a result of the pre-tender discussions – allowed for variations in the methodological approaches, realising that openness could be beneficial for the methodological match.

On the basis of the Terms of Reference requirements as outlined above, a multi level analysis was conducted. At the micro-level, the evaluating team employed a statistical analysis based on a survey – taking into consideration predecessor programmes and a mid to longer term time frame. This explored the innovation activity, the number of jobs and connection to public research institutions and – as far as possible – indirect effects. In addition, a range of case study was asked for to better understand appropriateness, efficiency and effects of the measures.

At the meso-level (technological areas, competence fields) the impact assessment explored the interdependencies of supported activities and the effectiveness of strategic actors and competence field managers. This took into consideration value chain aspects in order to understand the spill over of the support to third, non-supported parties and structural effects of co-operations and other interactions induced and supported by the programme. Originally the programme manager envisaged a quantitative, survey based control group approach; this was changed. The challenge of a control group approach, especially at competence field level, was to define and mobilise the control group. Even if a group of firms could be identified that showed the same characteristics as the firms that were supported, the incentives for them to participate are limited. Therefore, the control group approach employed an interview strategy, mobilising a smaller number of non-supported firms to discuss their perspective on the programme and also their development within the competence field and thus to compare trajectories of companies.

For the macro level, the effects for the economic system of the city-region, the results of the previous phases were to be extrapolated and complemented with an econometric simulation. This also put in place a manageable and reproducible indicator system that should enable a long term and sustainable monitoring of effects also in the future. The macro-economic evaluation has been conducted for the programme “support of research, innovation and technology” and the innovation assistant. A regional input-output model has been applied, based on the official statistics of the German Statistical Office and the data collected in the project. The main indicator used was *gross added value*. Gross added value is based on the products that resulted from the supported projects and the turnover they achieve plus consumption effects of jobs created. The effects were weighted with the likelihood that funded projects would not have been done without the support. Not included in this model are spill over effects to other, non funded companies and knock on-effects on scientific organisations, thus the macro-economic effect is underestimated. Further indicators for the macro-economic effects are *tax effects* (tax quotas in relation to the gross value added), and *job effects* (direct and follow on (indirect)). Finally, for those indicators the evaluation conducted a scenario analysis, whereby projections of developments of those indicators where altered according to different assumptions as to immediate effects. This resulted from the assumptions that effects could have been over-reported by the surveys (e.g. assuming turnover developments that were unrealistic). This allowed the programme owners alternative assessments of future

impact, inserting a conservative, "pessimistic" scenario and also made explicit how different assumptions lead to different projections of the gross added value in the future.

Finally, the Terms of Reference had asked for a complex analysis of the intervention by comparing it to the intervention portfolios and rationales in other regions and at Federal level. The comparison to other Federal States involved innovation and technology programmes of two neighbouring States. It did *not* include an evaluation of programmes in those States, e.g. could not compare effects and impacts. Rather it looked at strategies, foci, goals, instruments (mechanisms, designs), implementation processes and budgets, focusing on individual aspects that the evaluators (and the steering group) deemed important. The major idea of that comparison was to assess the context specific mix of measures in all three States to see if the basic orientation in Berlin fits the Berlin context and if not, if there are instruments and processes within them that could be transferred to the Berlin situation. This should lead to a comparative and complementary story, defining hampering and supporting framework conditions and programme characteristics. On that basis recommendations are asked for as to the future development of intervention logic and portfolio in the city-region, outlining an "optimal measure and financial mix" (ToR, p. 5) Finally, suggestions as to a better gender orientation were expected.

The programme management had asked for monthly meetings with the consultant. These were officially minuted, with clear action points for the following period. Thus, a strong monitoring of the whole process was established and results as well as methodological problems discussed almost in "real time". The meetings were organised with all important people involved from all the sub-programmes. All this helped to define the target definition and adjust the work programme.

This process is assessed to have been very valuable and successful. First, it allowed the programme management to learn about results early on and react as and if needed. Second, it simplified the sign off of the evaluation at the end. Third it enabled, as a side effect, a close, largely "de-politicised" interaction between different programme managers of different sub-programmes and any critical issues (such as analysing deadweight loss across different programmes, differences of small and large interventions, etc.) were resolved in those meetings. Fourth, it allowed programme specific adjustments to be discussed in the group. Fifth, and most importantly, it enabled learning across the board of management, about the breadth and interplay of the measures and about the various effects and how they can be demonstrated.

5. Effectiveness and efficiency of the evaluation

The cost – benefit ratio can be assessed as very positive, in the perception of the programme management the breadth and depth of the evaluation is high in relation to the costs of the evaluation. The consultants appeared to have performed more tasks than originally planned for (which might have been a result of the very close monitoring and interaction process in the course of the evaluation, as shown above). Further, the consultant was very good in communication in a broader sense, internally and externally, which then was extremely important to turn the evaluation into a marketing instrument, the presentation internally and at the press conference was regarded as a successful communication.

Overall, the methodology was sufficient and was able to achieve what was asked by the Terms of Reference. Two challenges in the overall design and ambition of the evaluation however remain. First, the management would like to understand better the macroeconomic approach and its results. While important lessons could be drawn from the macro-economic analysis, some of its methods were too sophisticated to be communicated in a way that was understandable for the informed layman, the technical complexity overwhelmed.

Second, the biggest challenge was the meso-level impact assessment. Meso levels have been defined as competence fields (certain technological areas). It was not possible to design and implement a quantitative meso level approach mainly because the available classifications of firms and product lines is not in line with the definition and reach of the competence areas (which are technological). A concordance between competence fields and economic sectors is missing, and thus it was not possible for the underlying statistical data to allow a sound assessment of impact of the programme on the competence field at large. All this would have been too costly to engineer within this evaluation, but it is a general problem for cluster and competence field programmes.

The survey was done very diligently and broadly, with a response rate of more than 50% response (including two reminders). However, the problem was that it is not possible to really establish what the funded firms represent in the competence field and what the intervention and its effects mean for the regional competence cluster as such. It would be essential to understand how the cluster as such has developed and what slice of the region the sample of funded actors represented. For different competence areas the situation was different, as according to the estimate of the evaluating team in some of the areas the coverage of firms in the programme and thus the survey was very high, e.g. results in that group were more accurate in reflecting the competence field more broadly.

The operational conduct of the survey faced one key challenge. The relevant contact data was not stored by the bank responsible nor did it fully monitor the contacts in firms and other organisations. Thus, in many cases it was hard or impossible to identify the right people, especially as information on early developments was needed. One extreme example concerned a large hospital with 45 project involvements, where it was hard to establish contacts for all the projects because of staff fluctuations. Further, to limit the burden on the firms, in one programme the evaluators had to rely on a survey already done in the process of the monitoring, while they would have preferred to conduct their own survey to tailor it better to the research questions.

In general, the survey and interview respondents were very responsive and invested a lot of time, despite the lack of a formal requirement for the funded firms to participate in an evaluation (which was subsequently changed: firms are now obliged to participate if they get funding). The readiness for interviews was very good and the programme manager did not interfere in the interview process.

There were no limitations in the process as it was all inclusive, the political decision maker was mobilised as were Parliament and neighbouring administrations.

One area under-evaluated was the relationship between the emerging networks in the city-region and the programme, whereby the programme might sometimes trigger networks or interact with them. It would have been of major importance to trace the development of competence field networks but, for the reasons mentioned above[1] (delineation of those networks), this was not possible. In the future, the programme management seeks to employ a network analysis.

A further area underdeveloped at competence field level is the relation of the cluster with its actual users: the interplay of the funding schemes and the clients and partners of supported companies outside the cluster are not being explored in the evaluation, which is a rather typical limitation of competence field or cluster level approaches, thus potentially underestimating the effects of the programme.

[1] The aforementioned networks are publicly funded GRW Networks (joint task "Improving regional economic structures"). The responsibility for this support network is in the Senate Department in another Section, so the interviewee was not directly involved in ad the evaluation. The promotion of technology transfer was not for the same reason included in the evaluation.

6. Results of the evaluation

The results of the evaluation were summarised on two levels, (1) overall, for the innovation and technology support represented by the measures evaluated, and (2) at the level of individual measures. Space does not allow elaborating on the details at both levels. The evaluation could not as yet analyse the mid to long term economic impact of the measures.

The overall set of measures in innovation and technology support in Berlin is assessed to have a high acceptance in the business and scientific community. The evaluation suggests that more sophistication needs to be applied in tailoring the financial conditions of individual measures to the kinds of projects and the market failure represented in those projects. The evaluation does not find any selection bias in funding, i.e. the programmes are open to all types of actors with all types of support need and the programmes appear to be flexible enough to accommodate for idiosyncracies. The only gap appears to be in funding of early stage firms. Further, the macroeconomic effect (assessed following the methodology outlined above), as far as could be judged already, were assessed positively while on the micro level there are no major deadweight losses to be seen.

In terms of governance, the cooperation with other Federal States has intensified, but is still obstructed by a range of hurdles. More cooperation between funding instruments across States and between State and Federal level is found to be needed.

On the level of individual instruments the results were, by and large, positive. No measure was assessed negatively in terms of its overall approach and effectiveness. In the following a short summary of the assessment is given, whereby the *result indicators used are in italics*.

Especially for the programme "upport of research, innovation and technology" the macroeconomic effects were found to be clearly beyond the costs of the measure (*gross added value, tax effects, job creation*). On the level of individual companies, the *readiness to innovate* has significantly increased through participating in the programme and it helped to mobilise third parties to invest in innovation in those companies. The future fond, despite a rather low programme budget and thus limited outreach, nevertheless succeeded in creating *added value* as technology platforms were funded which established *new innovation structures* in Berlin and as the *readiness for innovation* in firms has markedly increased. Problems relate to the *limited translation* of scientific knowledge into economic application and the reluctance of some companies to engage in larger networks. The programme innovation assistant succeeded in *supporting SME to hire experts* in certain areas, but it was *less successful in supporting knowledge absorption* from the public research base. Nevertheless, the macro- and microeconomic (*job creation, gross added value(turnover*) effects are assessed as being high, and the programme accelerates and improves *access of graduates into the economy*. The venture capital funds have performed well, with a *leverage effect of 17.7%* of the public investment from the funds. The reach of the measure in terms of *number of companies supported* is limited, but given the budget still appropriate. Long term effects could not yet be determined. The Technology Coaching Centres have shown limited thematic scope, focusing on business concepts and neglecting soft skills development. Within this scope they have *supported a range of new firms* and thus have contributed to the successful foundation of companies.

The comparison to other States found that there are a very similar set of measures with some considerable differences in designs and budgets, though. Those differences were explained by the different contexts of the comparator States, and the evaluation did not recommend a transfer of measures elsewhere, but rather confirmed that the basic orientation of the Berlin instruments is appropriate for the Berlin context.

Finally as for the synergy of the measures, the evaluation is very limited. It states that the "interplay of instruments as it was sketched out in the conceptual considerations of

the evaluation, can only be observed in a limited number of supported cases, but does not function in the breadth it could" (Final evaluation report, p. 125). The approach to measure synergies was to check for interplay of measures at the micro level, not to assess the overall macro effects of all measures combined.

The evaluation contains a whole range of recommendations both at programme level and as for the interplay of measures. At the overall level, to which we can limit the summary for the purpose of this report, the recommendations are grouped into four levels: improving personnel transfer, aiming at better synergies and interplay of the instruments along the innovation cycle (so that the individual measures better build upon each other), improving early phase financing, improving soft skill support, improving of gender balance and family friendly framework conditions, enhancing the involvement of proposers and funded actors in evaluation exercises.

7. Conclusions and lessons learned

This evaluation is a good example of a broad, thorough and multi-level evaluation. One characteristic appears to be the in-depth knowledge of the programme management about evaluation and about the effects of their programme(s) which is a very sound basis for the design of the Terms of Reference. Further, the differentiation between levels of effects simplified the analysis as it disentangled firm, competence field and macro-economic effects.

The very process of the ToR design has been fairly closed, but before the actual tender was designed, interaction with experts in the field who have an experience in evaluation led to crucial changes in the ToR, i.e. there was an adaptation as to how to tackle the macro-economic effects. Thus, the quantitative control group approach as originally envisaged by the programme management was abandoned and the tender text was formulated in a much more open way, allowing for methodological variation and an adjustment of methods to the skills of the consultant.

Restriction of the tendering process to the German market – below threshold – limited the competition, but – in the eyes of the management – increased the likelihood that those tendering are knowledgeable about the national and State level context.

One key success factor of the evaluation was the intensive consulting with the multi actor accompanying board of programme managers and owners. The consultations between the evaluation team and that board was intensive, but did not in any way limit the evaluation approach or interfere in a counterproductive way. While this could – at first sight – be seen as being mechanistic and a burden to the consultant, it proved to be crucial for the real time learning and feedback, so that both sides in effect were mutually re-assured that the activities performed went in the right direction and the challenges encountered could be tackled together, adjusting processes and methods as needed. Further, the involvement of all programme managing departments across a range of diverse programme proofed to be a very beneficial learning device.

In this specific case the interviews suggest that the relationship between the evaluator and the main programme owner and commissioner of the study was excellent and highly productive. The evaluating team was encouraged to be as critical as possible and was not limited in any way in their analysis. In principle, however, due to the dependency relationship between evaluating team and client – who might be a potential client in the future – one should consider to have neutral bodies detached from the programme management to commission evaluations.

As to the breadth of the evaluation, it proved extremely useful to evaluate a portfolio of measures not only at meta-level, but also as a combination of individual programme evaluations. EFRE (European Regional Development Fund) evaluations are of a similar nature, a number of inter-related measures have combined effects and thus simple top level or selected instrument level evaluations do not give the full picture.

As for the overall appropriateness of methods, the methods used in this evaluation were by and large very appropriate for the evaluation questions asked. The survey technique was implemented competently and robust and, as should be the case, complemented by case studies. This allows the combination of detecting the broader picture on the one hand and understanding causalities and the role of specific context conditions and programme requirements on the other hand. The envisaged control group approach would have given an even more complete picture, but given the challenges faced to construct a robust control group the decision not to do it rather than to do a slightly problematic one was the right one to take. Quantitative control group approaches need a very solid control group definition approach. The design of control groups in regional settings is challenging and needs to be designed in a very sound, time consuming process. Even if control groups can be designed, the readiness of the firms in the control group to participate can be expected to be low. The qualitative alternative employed in the evaluation can make up for that limitation to some extent, but does not deliver comparable, generalisable data to allow for a systematic comparison of developments between funded and non funded firms. The econometric modelling was relevant, but suffered from a lack of translation into lay terms, the practical usability was limited.

This leads to a more general limitation, i.e. the tension between the quest for a clear and simple number as to the overall effect, mainly as a result of the macro economic analysis. The press conference and reporting after the evaluation put some pressure on the evaluation team to focus on a simple set of numbers rather than report on the complexities of the programme. The macro-economic analysis delivered some numbers, but the focus on those figures in the wider discussion following the evaluation – not within the programme management or the political decision makers – was counterproductive.

A very simple, technical limitation is the lack of sound monitoring and contact data to conduct targeted surveys. Ideally regular monitoring should contribute fully to the evaluation exercise (and survey thus are limited) and contact database are updated regularly.

This example shows the importance of an intelligent programme owner who knows the importance of differentiating effects and can interact with the evaluating team on methods and approaches. The process to design and conduct an evaluation can be designed tightly without being a straightjacket to the evaluation process. On the contrary, it can enable learning within the administration.

With hindsight, the evaluation was too ambitious, and only the interaction with the external consultants led to adjustments of the process to make it more realistic.

Case study completed: October 2011