
MAURICE BASLÉ
University of Rennes 1, CREM-CNRS, Rennes, France. Email: maurice.basle@univ-rennes1.fr

(Received April 2005: in revised form December 2005)

BASLÉ M. (2006) Strengths and weaknesses of European Union policy evaluation methods: ex-post evaluation of Objective 2, 1994–99, Regional Studies 40, 225–235. The paper considers ex-post evaluation of Objective 2 at the regional level, focusing on the experience of France over 1994–99. These evaluations were ambitious, and when looking at the reports it identifies both strengths and weaknesses. The paper considers issues that are problematic in the case of the Structural Funds – such as adjustment lags, causality, measurement and ‘cross-checking’ of results using national data – which are explored both through a case study of an ex-post evaluation for the Brittany region, and through examination of the mid-term evaluations for France in 2003. While the paper does find some improvements in the quality of evaluation in the recent mid-term reports, it also identifies continuing problems. These include a weakness in the ‘logical diagram of impact’, tracing the chain of causality from actions to impacts, so that the paper argues for a more-detailed model of intervention.

Objective 2 Ex-post evaluation Regional development France


Objectif 2 Evaluation ex-post Âménagement du territoire France


Ziel 2 Ex-post-Evaluierung Regionalpolitik Frankreich

INTRODUCTION

The growth in the evaluation of public policies is a universal phenomenon. In the European Union (EU), the European Commission has been a pioneer in the new culture of accountability and assessment that is at the heart of the evaluation of socio-economic development. The evaluation of the European Structural Funds is a process (an application of procedural rationality) that produces outputs, such as a report on outcomes, process or quality, or some new knowledge about the efficacy, efficiency and impact of the Funds. The principal goal of evaluation is to collect information that can be directly used for improving public European action. In this sense, evaluation is not purely scientific research, and an evaluation report is not simply an academic matter. The evaluation challenge is to produce directly useful knowledge that is of value to society. As such, evaluation by the European Commission can be viewed as a form of participation in the European ‘knowledge-based society’ (Baslé and Renault, 2004).

Evaluation at the EU level is the result of a process that was initiated by European regulation, currently in its 1999 iteration (European Council, 1999, Articles 40–44). This regulation designed the process for allocating, budgeting, implementing, controlling, monitoring and evaluating Structural and Cohesion Funds. In each Member State, the managing authorities of Structural Funds programmes are obliged to adopt the European evaluation procedures at both regional and national levels. The European Commission also has regulatory obligations with respect to ex-post evaluations, and accordingly it carried out an ex-post evaluation of all Objective 2 programmes for the 1994–99 period, the results appearing in 2002–03. The evaluations form part of the overall Commission process of evaluation for accountability purposes and are also of internal management interest to DG REGIO (European Commission, 2004a). Additionally, they represent new knowledge for reorienting EU public policies.

The evaluation commissioning process is a huge undertaking, characterized by weaknesses typical of undertakings of this kind. The quality and utilization of evaluation are difficult issues, not least because it is well known (by those who consider the methodological approaches of public policy and study evaluation reports) that ultimately evaluation cannot open the ‘black box’ that would enable assessment of the real net impact of the Structural Funds (Tavistock Institute with GHK and IRS, 2003). This weakness is due to many factors, including the complexity of the policies; the lack of quantitative or qualitative data on policy monitoring; the difficulty of establishing a counterfactual (‘what would have happened otherwise?’); the attribution of impact to causes when many causes may intervene; problems associated with the ‘additionality’ of EU Cohesion policy (i.e. Structural Funds must be in addition to national, regional or local action); and the issue of objectivity when models or natural experiments are lacking.

These difficulties have been discussed extensively in the academic literature on Structural Funds (e.g. Bachtler and Turok, 1997; Davezies, 1999). They also have a contemporary political dimension, with the quality of impact assessment of Structural Funds being used to criticize the effectiveness of EU Cohesion policy in the Sapir Report (Sapir et al., 2003) as well as in Member State debates (HMT et al., 2003; Tarschys, 2003). In France, a report by the Commissariat Général du Plan (2003) was critical of the possibility of measuring the impact of the Structural Funds, arising from the so-called ‘sprinkling effect’, i.e. widespread actions and consequential effects – 26.9 million people have benefited from the Funds in France during the 1994–99 period, and an estimated 29.2 million people over 2000–06. Intrinsic difficulties arise in connection with the small size of some of the designated areas, the choice of areas that are not always well-described by statistical data and the long-term nature of some effects (Commissariat Général du Plan, 2003, p. 71).

Beyond the well-known imperfections of the scientific approach in human and applied social sciences, these failures are also a reflection of management and other problems, notably: the unavailability of appropriate statistics and data before the evaluation is commissioned; the difficulty of producing adequate terms of reference; the tension between the need for an historical evaluation (i.e. ex-post); and the need to prepare for future decisions, etc. Notwithstanding these weaknesses, the Structural Funds evaluation process is arguably one of the best-managed in the Commission. In particular, the quality of the results obtained...
enables their use in the process of reporting at three levels (European Parliament, European Council and European Court of Auditors). The results have also been useful to the European Commission, and seem to have been used to help design and implement the future EU Cohesion policy.

Objective 2 of the Structural Funds has been evaluated on several occasions over the past decade. An ex-post evaluation of the 1989–93 programmes was carried out (ERNST & YOUNG, 1997); many of the 1994–99 programmes were evaluated at the mid-term point, and all were evaluated ex-post (CSES, 2003); and the current 2000–06 programmes were most recently subject to mid-term evaluation (EUROPEAN COMMISSION, 2004c). The present paper considers, in particular, the ex-post evaluation of 1994–99 to illustrate the methodological challenges of evaluating the efficiency and effectiveness of complex regional development programmes.

The aim is to examine critically the strengths and weaknesses of the Objective 2 ex-post evaluation process. For the programming period 1994–99, the evaluation cycle was closed by a report presented to DG REGIO (CSES, 2003). The present paper examines the process of evaluation and results of the DG REGIO report. In addition, to extract lessons for future evaluation, an ex-post evaluation for a specific region of France (Brittany) is examined in some detail, and the results from mid-term evaluations of Objective 2 in France are also considered and recommendations made.

Overall, the paper emphasizes the near-impossibility of obtaining what evaluators in principle would seek to obtain, i.e. a reliable and credible assessment of impact, and a true legitimization for the EU interventions. This arises because of the context of multilevel governance, the ‘additionality’ issue and the heterogeneity of local actions and local areas. It is independent of procedure and practice in monitoring and evaluation. For the future (i.e. ex-post evaluation of 2000–06 programmes and beyond), it signifies that a new form of commissioning should be developed and that new forms of implementation should be introduced. Further, the monitoring process must be improved, so that there is a better foundation in knowledge from databases that are constructed at the regional level.

**EX-POST EVALUATION OF OBJECTIVE 2: A DUAL APPROACH**

Based on the evaluation report prepared by the Centre for Strategy and Evaluation Services for DG REGIO on the ex-post evaluation of 1994–99 Objective 2 programmes (CSES, 2003), four key issues related to the quality of Objective 2 ex-post evaluation are examined: political and strategic choices in relation to ‘needs’ and ‘context’ (i.e. priorities and areas); delivery methods; cost-effectiveness ratios (e.g. employment); and issues of ‘proportionality’ (i.e. relevance of the assistance compared with ‘need’); and economic impact.

**Outline of the ex-post Objective 2 evaluation report**

From the perspective of the new 2007–13 programming period, the ex-post evaluation of Objective 2 is of historical interest. What has been observed may be obsolete in relation to subsequent socio-economic changes, while future Objective 2 funds will be allocated to different priorities and areas. However, beyond the historical interest of an ex-post evaluation, the results of this analysis may still be useful for the future.

The CSES (2003) report considered the ‘call for tenders’, the evaluation ‘agenda’ and the ‘results’, and gave an assessment of their quality. With respect to the most important element of the evaluation, the impact assessment of Objective 2 interventions has two elements, as outlined by DG BUDGET in its ex-post evaluation report (EUROPEAN COMMISSION, 2004a).

First, the Structural Funds have an impact on the policy process. This is explained simply by the budgetary and implementation requirements of the Funds, which provide political and territorialized ‘impulses’ to relations between different levels of government. The report concludes that the ‘greatest added value of Objective 2 programs . . . arose from the development of powers and autonomy in the regions, thus strengthening local participation’ (EUROPEAN COMMISSION, 2004a, p. v). Second, with respect to economic and social impact assessment, the report concluded that:

> while the contribution of the program to job and wealth creation was judged as significant, the effectiveness of many measures was difficult to assess because of poor target-setting practices. In terms of job creation, the areas of R&D, innovation and technology transfer as well as business support measures performed especially well.

(EUROPEAN COMMISSION, 2004a, p. 21)

Notwithstanding these caveats, the conclusion of the ex-post evaluation (CSES, 2003) was that Objective 2 had a positive impact on competitiveness and cohesion in many policy areas (i.e. regional policy, policies for rural development, research and development (R&D) and information and communication technologies (ICT), as well as policies such as social, transport and energy policies). The following sections now examine the methodological basis for this conclusion in more detail.

**CSES report**

The CSES Report (2003) examined Objective 2 programmes with total EU funding of €16.5 billion allocated in 82 regions with 62 million inhabitants
(17% of the EU’s population). It used a twofold approach. First, a ‘top-down analysis’ of statistical indicators was used to identify impacts at European, national and regional levels; and second, a ‘bottom-up analysis’ was undertaken involving regional analysis to assess the ‘intervention logic’ and impact of Objective 2 programmes using micro-economic methods. The results from the top-down and bottom-up analyses are considered in turn.

**Top-down analysis**

The CSES Report began by examining impact using measures corresponding to the eligibility criteria for Objective 2 funding. The eligible regions were designated based on key indicators (unemployment rates, production structures and per-capita GDP), and the report examined the evolution of these indicators over 1994–99. It undertook a comparative-static analysis and benchmarking of these trends. The top-down analysis was not ideal, because in principle there are many alternative techniques that might be adopted: comparing development in a region that receives Objective 2 aid with non-assisted regions; comparing the before-and-after situation in a particular region that has received Objective 2 assistance; and a survey-based approach to obtain the views of informed observers and others on the extent to which Structural Funds intervention has made a difference (CSES, 2003).¹

Further difficulties arise because of a lack of consistency in the designation indicators used by Member States to define eligible areas and the differences between countries/regions in the starting points for development at 1994. Among the commonly used indicators were per-capita GDP (used everywhere except Denmark and Italy) and the unemployment rate. Economic activity was also used although the regions were heterogeneous (in 1993, 49.7% in the Belgian Objective 2 regions, and 66.9% in the Swedish Objective 2 regions). Other indicators were: industrial employment, although again there was heterogeneity (Voralberg in Austria had 45.9% of employment in industrial production, but Sweden had only 26.6%); net immigration, but these also differed widely (in 1994, 10.7% in Objective 2 regions in Luxembourg, but 1.9% in France); and proportion of the population educated to a medium or high level (this varied in Objective 2 regions from 80% in France to 39% in Germany).

Nevertheless, the CSES Report constructed a database of information on the 82 Objective 2 programme regions covering development indicators, the financial aspects of programmes and output data, and analysed it at national and regional levels.² A typology of regions was created using social and economic indicators for each year, 1994, 1997 and 1999. This involved creating ‘families’ of regions from the 82 Objective 2 regions based on the assumption that the impact of Structural Funds depends heavily on the structural context of a region, its initial conditions and the social adequacy of the programme. The typology proved simple, and three families of region were created. The first were ‘traditional industrial’ regions characterized by large manufacturing firms, high population densities and high economic activity rates, but relatively low levels of per-capita GDP. The second family was made up of ‘diversified economies based on urban centres’, combining industrial strength, urban decline and relatively fast growth in service activities (e.g. Berlin). The third was based on ‘small-scale industrial activities’, which had much lower population densities compared with the first two. France, with 19 Objective 2 regions, had 15 regions in the second category and four in the third family. Similarly, of four regions, Belgium had three in the second group and one other region in the final category.

**Bottom-up analysis**

In terms of a bottom-up analysis, the challenge was to determine the allocation of funds, both geographically and thematically.³ Questions of design and implementation were also considered.⁴ Consideration was given to potential downside factors arising from participation in Structural Funds, including costly administrative overheads associated with managing and delivering the Funds compared with purely domestic schemes, and the possibilities that regional effort or resource allocation might be ‘skewed’.

The main evaluation question was as follows: What progress did Objective 2 regions make during the 1994–99 period towards the aim of helping regions suffering from industrial decline to overcome structural problems and offer their inhabitants a quality of life comparable to the core parts of the EU? (CSES, 2003, p. 64). When considering this, it must be noted that a policy is first a ‘theory of action’, with expected effects arising from the utilization of resources. Only after implementation is the ex-post evaluation undertaken, but the quality of this will depend on the ability to analyse the evolution of indicators of eligibility, structural change and development (as impacts of the programme), such as unemployment or per-capita GDP. With Objective 2, the policy-makers do not explicitly detail the ‘theory of action’, but by examining European Commission documentation and laws, a broad ‘theory of action’ (or intervention logic) for Objective 2 can be extrapolated as follows:

- **Support for restructuring and stabilization (inputs)** leading to:
  - investment in business support, job creation, enhanced environment (outputs)
  - growth, wealth creation and recovery (outcomes).

The issue for the evaluator is twofold: comparison with an ex-ante benchmark and comparison across regions.

---

¹ CSES, 2003
² Baslé, 1999
³ Baslé, 2000
⁴ Baslé, 2003
As usual, the evaluator has to draw conclusions about efficacy, effectiveness or impact, and the efficiency of the aid. In so doing, there are a number of important issues highlighted by the CSES Report, some of which are especially problematic in the case of the Structural Funds interventions.

- Adjustment lags. The problem for the evaluator at the EU level, but also at a regional level, is that adjustments usually take time, so that measurement of these is based on expected effects, and not on controlled behavioural change. As the report stated:

> [a] process of this sort will take many years to accomplish, the development of a culture of entrepreneurship depends on wider attitudinal and societal changes, major physical infrastructure projects have long lead-times, there is a time-lag between improving the image of an area and an upward trend in inward investment, and a long-term investment in education and training is needed to develop a work force capable of supporting knowledge intensive economic activities.

(CSES, 2003, p. 66)

Thus, results presented in the CSES report could not be given with a high degree of precision, but which is no better than is usual for the results in terms of gross effects. They are expected gross impacts that involve a ‘fuzzy’ assessment of the correlation between aid and development, but in a number of cases, it is no more than that and could certainly not constitute a ‘proof’ of effect.

- Causality. As is widely known, ‘correlation is not causation’. Rigorously speaking, without structural models, natural experiments and counterfactual evidence, we do not know what would have happened if Objective 2 funds had not been allocated. Notwithstanding this, the report found evidence of convergence between the EU as a whole and Objective 2 regions on a number of key indicators over the period 1994–99. For example, ‘unemployment rates in the Objective 2 regions fell at a greater rate, and this trend was mirrored in the majority of countries’, while per-capita GDP ‘increased in the Objective 2 regions between 1994–99 from a level of 3.2 percentage points below the average at the beginning of the period to the EU average in 1999’ (CSES, 2003, p. 40). Other indicators, like migration rates and population change, also suggest convergence between Objective 2 regions and the EU as a whole, but the exception was educational attainment levels.

- Importance of outputs. When providing aid, the ‘best’ policy would seem to be one that has an obvious output, so that the intervention will not be necessary in the future. So, the observation that ‘areas with a predominance of small-scale activities … performed relatively well in terms of job creation’ (CSES, 2003, p. 157) was important for ‘proving’ the success of Objective 2 programmes. In the same spirit, the report observed that 300 000 small- and medium-sized enterprises (SMEs) received assistance, suggesting that this gave benefits to 3.6 million people, but without any further detail on the nature of the benefits. The Commission has an objective to stimulate SMEs and attaches great importance to this. However, the connection between a policy objective or financial input and an outcome is presented as a ‘given’, effectively creating a self-fulfilling prophecy.

- Measurement. The CSES report found evidence in past evaluations (1989–93) of gross jobs created, saved or relocated, and of the ‘cost per gross job’, which were projected forward to 1994–99 (taking inflation adjustment into account). There was no feasible way to cross-check the data, so that it can be inferred that these results are ‘expectations’, rather than scientific results from observation, tests or similar methods. The CSES report calculated that the 1994–99 projected gross impact could be 2 182 959 jobs (with a ‘cost per job’ of €7443), the net impact being 1.15 million jobs. However, using bottom-up analysis and monitoring data, the report found that ‘the estimate for jobs created was robust’, but that ‘the quality of monitoring data available from regional authorities meant that the estimates for jobs saved and jobs redistributed should be treated with caution’ (CSES, 2003, p. 138). One point stressed was a lack of differentiation between jobs saved or relocated. Further, a drawback with the data quality is that it derives from interviews with key partners, which has obvious drawbacks. More importantly, at the European level, the ‘definitions and methodologies varied considerably across different countries and regions’ (CSES, 2003, p. 139).

- Cross-checking. Finally, some imprecision was noticed in the results because it was infeasible to cross-check these with national data, a problem that presents a real barrier to better evaluation. Thus, the issue of ‘project additionality’ gets little more than a mention (i.e. activity that would have taken place without Structural Funds), while other issues, such as displacement (a loss of jobs elsewhere in non-assisted activities or areas) or the indirect effects (calculated by the use of local multipliers of expenses), are tackled only approximately. Hence, where the regions are large and there was little geographical fragmentation, the indirect effects were assumed to be contained within the eligible areas. Where this was not the case, it was assumed that spillover effects would have occurred on a more extensive scale with a lower level of indirect job creation in the assisted areas themselves (CSES, 2003, p. 145).
and Social Cohesion (European Commission, 2004b, p. xix), the Commission presents an assessment of quantitative impact:

Evaluation studies suggest that overall, structural intervention in Objective 2 areas led to the creation of some 700 thousand jobs over the period and just under 300 thousand in net terms, while around 300 thousand SMEs received assistance to improve their production methods and to seek out new markets. At the same time, some 115 million square metres of industrial waste land was cleaned up and reconverted . . .

However, this has not been subject to cross-checking and is subject to caution.

The paper now examines in some detail the results of ex-post evaluation at the level of a region. This enables consideration of the relevance of the methodology and utilization of evaluation for designing the new strategy for Objective 2 for 2000–06. The analysis is for the Brittany region, in north-west France, which has a population of 2.9 million.

The process of evaluation is monitored in France at the central level by the Délégation à l’Aménagement du Territoire (DATAR), and delegated to the Préfecture de Région Bretagne (the General Secretary for Regional Affairs). After a public tender, with a regional variation on the national terms of reference, a well-known consultancy firm was chosen to conduct the ex-post evaluation for Brittany. The report, published in March 2000 (Arthur Andersen, 2000), was conventional in nature, and the paper now considers some major features of this evaluation.

Implementation and delivery

As is typical for a regional evaluation, the Brittany study considered various evaluation goals: coherence, efficacy, efficiency, gross effectiveness, systemic and long-term impacts, utility, and sustainability of the programme. It was assumed that assessment according to each goal was feasible. The evaluation used a mix of tools (questionnaires, interviews, statistical analysis and review of the documents) in order to derive composite indicators for each goal. Each indicator had to provide a rating of one action from the point of view of each goal, e.g. ‘quite good’ for action X in the case of employment. The evaluation effort concentrated on high-cost projects that had been implemented (which was not necessarily the same as officially announced at the start of the project), and focusing on projects with the highest priority.9 The ‘standard’ Objective 2 programmes comprised a mix of support for enterprise, technological development, employment and infrastructure.

In Brittany, as in some other regions, the working assumption has been that the implementation of Structural Funds has not been successful. To evaluate this issue, an audit of the management and delivery of the Funds was carried out. This covered the project selection process, the role of partners in project selection, the understanding and acceptance of the process among beneficiaries, the monitoring of actions of both beneficiaries and eligible actors, and the ability of the organization to ‘induce success’. On these matters, the evaluation was standard, and the interviewees chosen were those directly responsible for delivery and control.10

In general, the orientation of the regional evaluation was towards implementation performance. However, the most difficult questions about evaluation concern the quality of impact assessment. For example, what is the model of impact assessment? Is it possible to succeed in impact assessment (or as a first step efficacy measurement) by using interviews and questionnaire surveys? These questions might seem overly ambitious, especially given the insufficient means and tools available at the regional level, but in this case it appears impossible to conduct a ‘bottom-up’ approach.

Logical diagram of impact

As is usually the case in regional ex-post evaluations, the Brittany report began by interpreting the official objectives and attempting to construct a ‘logical diagram of impact’. The aim of this diagram was to provide a framework to describe the expected chain of causality running from actions to operations, then from operations to outcomes and then to impacts both at a geographical level and over time. One of the principal difficulties was how to identify the key factors in a programme’s success, or lack of it. The real world is complex, not causal, and in essence the logical diagram is a ‘fiction’. This is because a logical diagram presents interventions as a causal chain, but the chain is hypothetical, and tests must be made to determine the existence of the relationships.11 Notwithstanding this, ‘a logical diagram of impact’ can be a realistic representation of reality, but only after a range of tests, such as statistical or econometric tests, the use of cross-checking or after collecting substantial qualitative material or documentation.

In the case of Brittany, the evaluation report presented a rigorous definition of the ‘logical diagram of impact’ and a summary of its potential use. Objective 2 was designed to support economic development through a range of measures (e.g. training and employment, tourism development, and environmental management), and a diagram was constructed for each of these, including key actors. The detailed procedure of implementation for each measure was outlined and questions of incentives and competency were addressed. The absence of strict administrative procedures was judged as signalling a general lack of coherence and efficacy. It led to the formulation of recommendations for better public management, suggesting a general failure in this area. However, the report did not conclude at
this stage, but it went on to produce a quantitative impact assessment.

Impact

Notwithstanding the relatively poor response rate to the questionnaires (36%), the report presented results and some general indications of the programme’s impact. Perhaps greatest confidence can be attached to those indicators for training support and aid for physical and non-physical investments. Drawing inferences from the questionnaires and interviews, the report presented detailed results for employment creation, which amounted to 2326 jobs in 491 projects, and led to a projection of a hypothetical 8082 jobs for the whole of Brittany as a consequence of Objective 2. Knowing the costs of the measures, the report deduced a cost-effectiveness ratio of FF62 000 (francs) for each new job created. However, there was no calibration of the results, i.e. no cross-checking with the stock of employment in Brittany, and no comparison with estimates for other French regions.

Experience tells us that, except for a few regions that use macro-economic models or econometric tests, evaluation practice is similar in many regions. Thus, for one (and probably many) regions, imprecision exists due to weak calibration at the regional level, which leads to errors if aggregation is carried out. However, even for a good regional evaluation report, it is likely that the data necessary for generalization to an aggregate level are not of a sufficient quality. As a further point, beyond the measurement of impact, the Brittany evaluation contains indications that the beneficiaries were ‘well satisfied’ overall. However, this is not an assessment of quality, coherence or efficiency, since, given rent-seeking behaviour, the Structural Funds will generally be well-received, like any other ‘gift’. In terms of impact, this apparent satisfaction does not offer anything more than a subjective account of what happened.

Overview

To summarize, the inadequate nature of evaluation ‘proof’ is apparent in many cases, and, with the exception of a minority of regions with the capacity for cross-checking, it is not possible to be overcome. In a number of regional evaluations, the ambition is clear and convincing, but when implementing the methodology, the evaluators chose a restricted number of tools and techniques and could not obtain representative data that can be validated by cross-checking. As such, the evaluation reports are well-informed documents about the evolution of regional socio-economic development, but, strictly speaking, they do not constitute impact assessment or a study of programme effectiveness.

Although the impact of Objective 2 is at stake, from some evaluation reports it is really only possible to infer the possibility of a weak change of trajectory of the regional economy. This is based more on the good performance management of Structural Funds and intuition, rather than on ‘logical diagram impact’ testing. For Brittany, it is the case that it experienced some demographic and socio-economic changes, and that the implementation of Structural Funds projects was an ‘impulse’ for growth, but it is not possible to infer causality between greater regional ‘dynamism’ and the operation of Objective 2. At the subregional level, the ex-post evaluation for Brittany found that there were differences of opinion, but these were probably due to a greater heterogeneity of context and to conflict about the credibility of ‘proof’ with socio-economic models, the logical diagram impacts, etc. The evaluation concluded that Structural Funds support was too small and had too weak an influence on infrastructure, while the report may be too optimistic on the high level of job creation. Thus, overall, it can be concluded that Objective 2 provides an opportunity for eligible regions to compete and win some funding. This creates an ‘atmosphere’ for development, but the effects triggered do not necessarily add up to an ‘impact’.

MID-TERM EVALUATION OF OBJECTIVE 2

The above views of Objective 2 ex-post evaluation are pessimistic, but do they also hold true for the mid-term evaluations of Objective 2 for 2000–06? For this, 20 of the 26 mid-term evaluations produced for France in 2003 were examined, and many of the same strengths and weaknesses were found. There was little cross-checking with macro-regional models or with other types of approach; there was an emphasis on implementation; many of the same diagnostic tools were used; and significant qualifications were attached to the potential or real gross effect of funds at a local or regional level. The following section presents some of the major lessons from the Objective 2 mid-term evaluations and considers some of the specific difficulties in obtaining good impact assessment in this case.

Lessons

First, an examination of the mid-term evaluation reports reveals a high quality of implementation and much better practice in the monitoring and evaluation of the Structural Funds at both regional and national levels (compared with the 1994–99 period). Overall, real improvement was observed in the average quality of reports since the ex-post evaluations, and excellent quality in some cases. Differences existed between regions, probably due to the competencies of the evaluators and the facilities provided by the commissioning
agencies of the evaluation. However, in general, there was a better mastering of the process of Structural Funds evaluation by the official authorities in the French regions. The complexity of evaluation has been partly overcome by a good division of labour between DATAR, the French national development agency, and the Secretaries for Regional Affairs, and better training for local government officials. Through ‘learning by doing’, French authorities have succeeded in the division of tasks and in the implementation of monitoring, quantitative collection of materials through standard software, and by enhancing competencies at different levels.

Second, there was a real double challenge for mid-term evaluations. They were attempting to assess in too short a time-scale into the implementation of the Structural Funds; the period between the start of programme implementation and the mid-term evaluation was only about 18 months. Rather than evaluation, the task undertaken by evaluators was more akin to simple monitoring. This indicated that effective mid-term evaluation is feasible only if data collection is recurrent and if software can record applications, awards and project activities.

Third, from the examination of the French mid-term evaluation reports, it is apparent that there are new knowledge management techniques for the monitoring and evaluation of the Structural Funds. The national agency DATAR has invested in relatively new software for the management of the Funds (known as PRESAGE) and this has been made available to the 26 regions, enabling network implementation of the regional management of Structural Funds.¹³ PRESAGE is compatible and universal, providing a coherent picture of the whole system of Structural Funds in France. In the 2007–13 programme period, there will be fresh challenges of dissemination and partnership, and to some extent the advantages of the new software will be lessened by the difficulties in obtaining good information that is homogenous, and so enabling the creation of good indicators.

Finally, it appears that the evaluators often lacked an academic or scientific critical perspective. This is particularly evident in cases where evaluators used qualitative answers from interviews that seem similar to a ‘conversation’ with respondents. Further, it can be observed that ‘benchmarking’ between similar or quasi-similar regions was rare (even though it would offer something like a counterfactual situation). In France, a lack of coordination between DATAR and the National Institute of Statistics (INSEE) is also noticed, despite a contractual relationship between these agencies.

**IMPLICATIONS AND RECOMMENDATIONS**

The process of ex-post evaluation of Objective 2 reveals the usual weaknesses of evaluation where contexts are heterogeneous, where the areas are tightly drawn, where ‘additionality’ causes problems of impact identification, and where there is a lack of formal knowledge about these contexts. The evaluation process of the European Commission is well regulated, and great progress has been made in terms of quality, so that the utilization of these results is appropriate. Of course, it may be that a biased view is presented, but Objective 2 evaluation seems a more difficult task than the evaluation of Objective 1. A problem is that the distribution of Structural Funds in Objective 2 regions is not as predetermined as is the case for Objective 1; while it supports national regional development and restructuring programmes, it is generally not the main component or the deciding factor in socio-economic change. Further, ‘additionality’ seems to be more of an issue for Objective 2, while its impacts are probably smaller due to a ‘sprinkling effect’, while contexts, path dependencies and the nature of regional integration are likely to be factors in determining outcomes.

The possibility of change brought about by the ‘announcement effect’ of Structural Funds eligibility is credible. A good ‘ambience’ surrounding the announcement of a Structural Funds award may have an impact, but while the ‘dynamism’ associated with the adoption of a project culture, and the prospect of Structural Funds benefits, may exist, it is difficult to translate these into a quantitative impact. As such, it is necessary to rely on the ‘implementation effect’. Some are doubtful about the size of this impact on competitiveness and employment. Further, Objective 2 funds are given to stimulate investment, competitiveness and employment in regions where structural adjustment is already in operation. A lesson from the analysis is that there are weaknesses in the past ‘logical diagrams of impact’, and a more detailed model of intervention should be designed and tested.

This will probably be easier in the future due to new developments in regional science. We now know more about critical factors of endogenous development, and it is something that can reinforce the efficacy, efficiency and effectiveness of Objective 2 interventions. At present, the weakness in evaluation probably has less to do with a general theory of action than insufficient explanations of the implicit reference model used by the European Commission, and which is in need of much greater articulation. Strictly speaking, the European Commission does not have an explicit model of territorial action, so that the impacts are assumed inevitably to take place. This is more a matter of deductive beliefs than of an empirically based theory of action. Thus, the issue is not about the quality of evaluation, as it could probably not be better given the lack of a reference model and of previous knowledge resources.

In future, more detailed modelling and testing should induce more detailed ‘logical diagrams of impact’ (e.g. Fig. 1), and more data should induce
more tests and hence more sophistication in the evaluation agenda and results. Before concluding, some items for this agenda can be suggested. In future, models of regional endogenous growth need to be developed for renewed legitimization of Objective 2. They will be based on the new hypotheses, but should have the following features.

Given that Objective 2 assistance is more limited than in Objective 1 regions, the future Objective 2 (Regional Competitiveness and Employment) has to be seen as compensation for under-performing regions, but only for those that can prove their ‘dynamism’. Every region has to reach its proper ‘efficiency frontier’. Performance has to be achieved within the constraint of ‘social acceptability’. It must allow for the possibility of simulation, if not the possibility of testing. In principle, it would demonstrate greater cohesion from more cooperation and more use of Structural Funds in a new style of Objective 2.14

In addition to mechanisms normally used when considering impact multipliers of external assistance on an area with differing characteristics (e.g. industrial attractiveness, local integration and capacity of reinvestment), the model has to be enriched by the inclusion of ‘innovation’ in its broadest sense.15

CONCLUSIONS

Evaluation of Objective 2 programme for 1994–99 was not a simple task. Indeed, controversy about the quality of assessment and impact are still live issues. There are intrinsic difficulties in evaluation, such as the small size of territories and the ‘additionality’ of funds, but there are also difficulties that could be overcome if some new developments were to emerge. These include a better ‘logical diagram of impact’ that incorporates new ideas from regional science models and better regional quantitative and qualitative data from

---

Fig. 1. Logical diagram of programme effects
better management and monitoring. Data and knowledge on impact also have to be produced.

In the near future, the European Commission will generate the new eligibility criteria for the post-2006 Objective 2, taking into account endogenous growth theory. Priority will be given in Structural Funds interventions to indicators of employment effects, including a better-quality calibration process of local multiplier effects, leading to a better distribution of funds, and better control of implementation, offering the greater prospect of obtaining the hoped-for effects. Thus, one can expect that with the lessons from the ex-post evaluations, the next round of Structural Funds for ‘Regional Competitiveness and Employment’ should be more effective. However, the net gain may be too weak to convince the more pessimistic. This is because the new approach to implementation must be accompanied by the following:

- Intellectual investment in the regionalization of macro-economic models.
- Working through the socio-economic effects at the regional level through a ‘logical diagram of impact’.
- Greater knowledge of the key factors determining success or failure.
- Investment in software for monitoring the costs, activities and effects.

Being better designed and controlled, the policy would become more attractive. Indeed, rather than concentrating resources on the poorest areas using Objective 1 only, in our view a more reasonable choice would be a combination of Objectives 1 and 2. In this case, Objective 2 should offer support to more attractive areas. While this runs counter to the social solidarity of the current Objective 2, the fact is that many regions are unattractive to investment. By supporting more attractive regions, the multiplier and spillover effects may be favourable to neighbouring regions that are poorer or in transition.

The final evaluation for 2000–06 and future ex-post evaluations will hopefully be more accurate and offer more fertile material for analysis. It is to be hoped that with 2007–13 in view, the implicit modelling corresponding to the Lisbon and the Gothenburg Agendas will be more fully articulated and developed at regional and national levels, with the setting of targets and time limits. It is also to be hoped that the new pragmatic evaluation culture will permeate every level of administration. The new Evaluation Guide from DG REGIO (TAVISTOCK INSTITUTE with GHK and IRS, 2003), while not free of digressions, has to be of interest to policy-makers and evaluators (Baslé, 2004; EUROPEAN COMMISSION, 2004a). A consequence of better European policy design, monitoring, evaluation and the utilization of evaluation could be a greater credibility for Cohesion policy in the EU-25 or EU-27. Policy-makers are in a new world, with a greater heterogeneity of regions, greater expectations from the new Member States and greater opportunism from some existing Member States, and the need for knowledge on the impact of European policies has probably never been greater.

Acknowledgements – The author is indebted to the referees and journal editors for valuable assistance in finalizing the latest version of the paper. He is also grateful to the Centre for Strategy and Evaluation Services for providing the Objective 2 evaluation report prepared for the European Commission. All remaining errors and imprecision remain the responsibility of the author.

NOTES

1. The merits and drawbacks of each technique was examined (e.g. problems of comparators, issues of performance for programme management and the risk of response bias).
2. Using exchange rates at purchasing power parity for GDP in the former case.
3. Where were the European funds allocated and for what kinds of programme or project? There were also questions of absorption capacity for high (or higher) levels of funding, and whether the local effect could be estimated.
4. For example, are the priorities and measures standardized? How well do the European Commission, Member States and regions coordinate their functions? And how well is good practice disseminated?
5. The latter is measured in purchasing power parity.
6. This is the third of the threefold typology of regions given above. They received 11% of funds and generated 13% of the jobs.
7. As opposed to the additivity of the Funds (i.e. which should not substitute for national funding).
8. At the start of 2006, DATAR was reorganized and renamed ‘Delegation Interministerielle à l’Aménagement et à la Compétitivité des Territoires’ (DIACT).
9. This seems reasonable, but there was no analysis of impact of widely disseminated ‘soft’ projects.
10. The questionnaire was sent to all 1963 beneficiaries, yielding 714 responses.
11. In detail, the chain consists of interventions, leading to activities and operations, followed by an implementation process, intermediate results and final outcomes in the dimension of goals (official objectives) and emergent effects (systemic impact).
12. For example, a satisfaction index of (say) 85% could indicate many things, e.g. satisfaction with the presence of aid, the real satisfaction of a need, the pleasure from a windfall profit, etc.
13. For details, see http://www.datar.gouv.fr/datar_site/datar_europesom.msf
14. For consideration of these kinds of virtuous circle, see SUIRE (2005) and BOSCHMA (2005), for example, on interactive and creative clusters.
15. More innovative regions have the capacity to transfer technology, provide coherent training for workers, network externalities, local capture and provide a ‘reputational’ feedback mechanism.
17. A good example is the official French software for Structural Funds, PRESAGE, noted above.
REFERENCES


