THEORETICAL PERSPECTIVES ON APPROACHES TO POLICY EVALUATION IN THE EU: THE CASE OF COHESION POLICY

JULIAN HOERNER AND PAUL STEPHENSON

Cohesion Policy accounts for the largest area of expenditure in the EU budget. Because of its scope and redistributive nature, evaluation is particularly important. Policy analysis tends to overlook the evaluation stage. Few empirical studies seek to apply theory to EU policy evaluation. This article questions the relevance and usefulness of theorizing evaluation practice, exploring positivist, realist, and constructivist perspectives upon approaches to evaluating Structural Funds Programmes. It illustrates how political science theories can provide scholars with useful insights into the way EU policy evaluation is carried out. It develops a toolkit for analyzing real-world approaches to evaluation and then applies it to three separate Cohesion Policy programmes. The analysis shows how, from a theoretical perspective – and contrary to the mixed methods rhetoric of the European Commission – positivism remains the dominant approach when evaluating the Structural Funds and considers why this is so, identifying the ability to demonstrate efficiency and effectiveness, cost, influence, and evaluation culture as key characteristics.

INTRODUCTION

In early 2010, European Union (EU) Cohesion Policy overtook the Common Agricultural Policy (CAP) for the first time as the biggest area of expenditure in the EU budget. Because of its large thematic scope and geographical scale, evaluation is of particular importance, and as such, considerable financial resources are dedicated to it. However, compared to the CAP – a redistributive policy with a similar value – Cohesion Policy is subject to more evaluation activity, arguably because it is particularly politically sensitive at the national and regional level, with a widening gulf between net donors and recipients. Since the 2000–06 programming period, and in the context of enlargement, there has been a significant increase in the quantity of evaluation activity, and moreover, a greater focus on methodology and different approaches to evaluation by practitioners and the scholarly community.

In general, evaluation in the EU is an under-researched topic. Even though evaluation at the EU level was introduced earlier than in many member states, and for a considerable number of policies, compared to other stages of the policy cycle, academic analysis of evaluation remains rather scarce. Indeed, where evaluation is analyzed, there is little theorizing in terms of reflecting upon the factors that influence how evaluation is carried out, and the patterns that can be observed. In other words, there is an abundance of evaluation reports and documents, from academic and commercial sources, but little abstract thinking about evaluation as a stage in the policy cycle. Academic analysis of evaluation by EU politics/policy scholars remains rather scarce.

In times of large fiscal deficits and tight government budgets across much of Europe, research is valuable and timely. This article’s aims are threefold. First, assuming that positivism is dominant in approaches to evaluation – such an assumption at least

Julian Hoerner is a Research Student at the European Institute, London School of Economics and Political Science, London, UK. Paul Stephenson is in the Department of Political Science at Maastricht University, Maastricht, The Netherlands.

Public Administration Vol. 90, No. 3, 2012 (699–715)
© 2012 Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.
commonplace in evaluation circles – it investigates what dominance actually means for
evaluation practice. Second, it examines the gap between the rhetoric on evaluation from
the Commission and what can be observed in practice – what is striking is the gap between
politics and real-life action. In the EVALSED guide, the Commission advocates that ‘it is
best to apply methods and techniques in combination as part of every particular evaluation
assignment’ (European Commission 2008, p. 97). However, in practice it expects results to
be mainly of a quantitative nature (Bauer 2006b, p. 109). Thus, it at least implicitly favours
a positivist approach. Third, the article systematically compares different approaches to
evaluation through a micro-level analysis, attempting to theorize evaluation in a similar
way to which theorizing occurs at other stages of the policy cycle. In so doing, it is mainly
explorative and based on a relatively small number of cases, its main added value being
to provide a new, transferable conceptual framework for comparing different theoretical
approaches to evaluation, rather than providing an exhaustive account of EU evaluation.
Considerable empirically-grounded work on evaluation practice has been undertaken,
such as in the 2006 special issue of Regional Studies edited by Bachtler and Wren (2006),
and by Molle in his seminal book on Cohesion Policy (Molle 2008). The theorizing of
evaluation is essential since, in contrast to other stages, our understanding of theory’s role
in EU policy evaluation is only weakly developed.

This article first provides the background to Cohesion Policy and the history of
evaluation in the EU. It then puts forward a framework, identifying the five most
important characteristics found in the literature regarding approaches to evaluation, and
thereafter applies this framework to investigate the different evaluation methods used
in Cohesion Policy for the mid-term evaluations in the 2000–06 programming period. In
so doing, it compares the evaluation of three programmes implemented at the national,
regional, and local level, namely the Objective 1 and 2 programmes and a Community
Initiative (Urban II), all wholly- or partly-funded by the Structural Funds (SF). Thus, a
small-N case study design is set up to analyse how the different theoretical approaches
are used in practice.

The methodology is a textual analysis of official programme documents produced for the
mid-term evaluations, chosen for reasons of topicality, timeliness, and comprehensiveness.
These evaluation reports provide detailed insight into achievements halfway through the
programming period, and also identified gaps still to be filled. The ex-post evaluations
of the three programmes were not available at the time of research and there are no
mid-term evaluations in the period 2007–13. Conducting an evaluation can take 12 to
18 months. The Commission’s own ex-post evaluation of Cohesion Policy programmes
for 2000–06 was only published in April 2010. The mid-term evaluations were significant
in the evolution of EU policy evaluation, though have recently been replaced by newer
instruments, as discussed in the following section.

A set of interlinked research questions is used. First, how do we theorize different
approaches to the evaluation of EU Cohesion Policy, and is political science theory
relevant? Second, if so, is one theory (positivism) dominant when evaluating Cohesion
Policy programmes, and if so why? Exploring these questions should improve our
understanding of theory’s role in the evaluation, but also provide new insight into ‘real
world’ activity in the policy process.
EVALUATING EUROPEAN UNION STRUCTURAL FUNDS

Cohesion Policy

Cohesion has been a political, economic, and social goal throughout the history of European integration (Leonardi 2005, p. 217). Cohesion policy is a massive area of budgetary expenditure, amounting to 225 billion euros in 2002–06, and has since increased to 347 billion euros in 2007–13, as agreed at the European Council in December 2005, i.e. one third of the total EU budget (European Commission 2011). Since the reform of the Structural Funds in 1988, and creation of a Cohesion Fund at Maastricht (1993) – complementing the European Social Fund (ESF) and European Regional Development Fund (ERDF) – the policy’s overall aim has been ‘the reduction of regional and social disparities in the European Union’ (Hooghe 1996, p. 3). However, the effectiveness of EU Cohesion Policy has been strongly criticized, especially by some richer member states. Begg (2010) calls Cohesion Policy ‘a policy searching for objectives’, and emphasizes the confusion created by different, often competing yet incompatible policy goals.

Prior to the 2007–13 programming period, its instruments comprised three mainstream ‘Objectives’ (1, 2, 5b) plus several Community Initiatives (Urban, Interreg, Leader, Rural). Objective 1 (renamed ‘Convergence’) is the programme aimed at those regions with a GDP of less than 75 per cent of the Community’s average to which the largest part of Cohesion spending was allocated in the 2000–06 period. Objective 2 (renamed ‘Competitiveness and Employment’) was concerned with revitalizing areas suffering from industrial decline and high unemployment (Hooghe 1996, p. 35). Various Community Initiatives (now captured under the third strand, ‘Cooperation’) funded smaller cooperation programmes in thematic areas not covered by the mainstream Structural Funds. One such initiative, URBAN II, had a relatively ‘modest’ budget of 72.3 million euros in the last programming period, its predominant focus being the economic and social regeneration of urban areas suffering from industrial decline.

Evaluation in the European Union

Evaluation became important for assessing the Structural Funds in the 1980s, given the redistributive nature of what was then termed ‘Regional Policy’ (Summa and Toulemonde 2002, p. 409). Most evaluation studies were carried out by external bodies and not by the Commission departments themselves, owing to a perceived need to ensure objectivity and neutrality, as well as a clear lack of human resources in the Commission Directorate-Generals (DGs) (Summa and Toulemonde 2002, p. 419). Today, Cohesion Policy evaluation is important not only for national, regional, and local authorities to improve implementation processes, but also for the Commission, in order to improve regulation (Leonardi 2005, p. 74).

In the latest policy reform, the member states were given greater obligations at the evaluation stage – they must designate an audit authority to oversee implementation and management, and take responsibility for detecting fraud (European Council 2006; Manzella and Mendez 2009, p. 20). However, the Commission oversees the work of these authorities and can carry out ‘on-the-spot audits’ at short notice (European Council 2006). Moreover, the Court of Auditors oversees the legal and technical aspects of Structural Funds spending. Currently, a new reform of Cohesion Policy is underway, which will probably change the way evaluation is carried out. The Commission has proposed introducing a so-called ‘performance reserve’, whereby money would be taken away from regions lagging behind in implementation (European Commission 2010, p. 2). Clearly, this would further increase the importance of evaluation as a basis for decision-making.
EU Cohesion Policy may change in terms of thematic priorities and even functioning, but evaluation will continue to have serious political implications. An evaluation typically answers questions concerned with: first, relevance (do the objectives correspond to the needs and problems?); second, effectiveness (to what extent were the set objectives attained?); and third, efficiency/cost-effectiveness (were the results achieved at a reasonable cost?). The aim is to provide a reliable and objective assessment that is accessible to all stakeholders. Since 2010 the Commission also uses ‘fitness checks’ – comprehensive policy evaluations assessing whether the regulatory framework for a policy sector is fit for purpose. They are meant to identify excessive administrative burdens, overlaps, gaps, inconsistencies, and/or obsolete measures which may have built up over time, often owing to the cumulative impact of legislation.

Practically-speaking, programme evaluations have traditionally taken place at three different points in the programme cycle of the Structural Funds: ex-ante evaluations to assess potential problems and verify targets; mid-term evaluations to assess progress and undertake changes during the programming period; and ex-post evaluations to assess the success of past programmes and to draw conclusions for future programming periods. As mentioned above, for the 2007–13 programming period, mid-term evaluations were made optional and largely replaced by an ongoing review mechanism that gives more flexibility to the member states (Manzella and Mendez 2009, p. 20).

TOWARDS A TOOLKIT FOR ANALYZING APPROACHES TO EVALUATION

Characteristics of different approaches to policy/programme evaluation

A comparative analysis of the characteristics of evaluation may reveal similarities and differences in approach between, or even within, policy areas. A review of the evaluation literature reveals specific concerns for the evaluator – precisely those characteristics that may determine how evaluation is practised, and how practice (approach) influences findings (outcomes). First, let us consider the underlying epistemological assumptions and thus, the evaluator’s own philosophical approach to knowledge acquisition and the social-scientific task of evaluation. This ultimately influences whether or not he/she is seeking to produce objective, generalizable results of universal value, or else acknowledges the existence of intervening factors – such as the unique nature by which every individual views the world – and which do not allow for generalizations. Such an important distinction mirrors the classical debate in the philosophy of science between logical empiricism and more interpretivist theories (Mark et al. 2000, p. 141). Indeed, different epistemological assumptions lie at the very core of the respective approaches; all other characteristics essentially stem from these fundamental perspectives. Second, one must pay attention to the research methods used. A basic distinction can be made between quantitative methods – such as macroeconomic models, input–output analyses, statistics, and control groups – and qualitative methods, such as surveys of beneficiaries, stakeholder interviews, and case studies (Bachtler and Wren 2006, p. 147). Third, one needs to consider the direction of the evaluation research, which may determine whether data is gathered on an aggregate level from secondary sources, such as national public administrations and/or regional statistical agencies (Armstrong and Wells 2006 p. 264), or is collected directly from the stakeholders involved, in order to gauge experiences and opinions at the local level. Fourth, attention should be paid to the scope of stakeholder involvement, and whether evaluators only engage those directly participating in the programme or aspire to secure a more comprehensive, inclusive involvement of the community, however marginally
affected by the programme (Guba and Lincoln 2001, p. 2). Choices will be made with time and financial constraints influencing the breadth or narrowness of the study. Finally, the role of the evaluator in the evaluation process must be considered, i.e. whether he/she is a neutral, impartial observer, or considered an independent variable that influences the findings. Some take a ‘God’s Eye View’, as Putnam formulated it (Mark et al. 2000, p. 144), while others consider the evaluator a policy stakeholder in their own right since he/she is ultimately a recipient of technical assistance budgets, and an agent of the principal who pays for the evaluation (Guba and Lincoln 2001, p. 8). In the case of the mid-term evaluations, the managing authorities at the national and regional levels commissioned the evaluations and were thus the principals, the Commission being the ‘final beneficiary’ of the reports.

These different characteristics are summarized in table 1. Tables 2 and 3 build on this schematic summary, depicting how these characteristics are expressed from the different theoretical perspectives, and how they are mirrored in actual evaluation research. This allows the reader to crosscheck ‘at-a-glance’ the five characteristics at the conceptual, theoretical, and empirical level. Such a framework is not available to date and could be applied in the future to a broader research project.

### Theoretical approaches to evaluation

The three theoretical approaches explained in this section are the standard approaches used in evaluation research. Indeed, they are mentioned by the Commission in its EVALSED Guide (European Commission 2008, pp. 20–1) and in key scholarly works on policy evaluation (e.g. Armstrong and Wells 2006; Bachtler and Wren 2006). However, these theories are in practice rather broad, and there are significant variations or strands within each. For example, in other political science contexts, positivism might be considered an epistemology (a way of thinking about knowledge) (Blackburn 2005, p.163), and constructivism an ontology (a way of thinking about the nature of the world) (Blackburn 2005, p. 261). However, in evaluation research they represent two distinct approaches, which differ in both regards. In general, these theories are not considered in the same way in public policy as in international relations (IR).

---

**TABLE 1 Characteristics determining the practical approach to policy evaluation**

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Definition</th>
<th>Further references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epistemological</td>
<td>Is the result seen as objective and generalizable or is the importance of</td>
<td>Mark et al. (2000, p. 141)</td>
</tr>
<tr>
<td>assumptions</td>
<td>the socio-economical context and intervening factors acknowledged?</td>
<td></td>
</tr>
<tr>
<td>Methods/data</td>
<td>Are qualitative or quantitative methods used? Are data collected?</td>
<td>Bachtler and Wren (2006, p. 264)</td>
</tr>
<tr>
<td>collection</td>
<td>econometric models, interviews, surveys etc.</td>
<td></td>
</tr>
<tr>
<td>Top-down/bottom-up</td>
<td>Are large meta-level, or small, in-depth studies conducted? Are formal</td>
<td>Armstrong and Wells (2006, p. 264)</td>
</tr>
<tr>
<td>approach</td>
<td>questions set at the top or does loose enquiry begin at the grassroots</td>
<td></td>
</tr>
<tr>
<td></td>
<td>level?</td>
<td></td>
</tr>
<tr>
<td>Stakeholder involvement</td>
<td>Is a wide range of stakeholders (SMEs, citizens, local government etc.)</td>
<td>Guba and Lincoln (2001, p. 2)</td>
</tr>
<tr>
<td></td>
<td>involved or only a very small number of large entities?</td>
<td></td>
</tr>
<tr>
<td>Role of the evaluator</td>
<td>How are the evaluators themselves conceptualized? Are they seen as neutral</td>
<td>Mark et al. (2000, p. 144); Guba and</td>
</tr>
<tr>
<td></td>
<td>or as an independent factor in the evaluation process?</td>
<td>Lincoln (2001, p. 8)</td>
</tr>
</tbody>
</table>

---

Public Administration Vol. 90, No. 3, 2012 (699–715)  
© 2012 Blackwell Publishing Ltd.
Positivism
According to the classical positivist approach, it is possible to obtain objective knowledge through observation (Bachtler and Wren 2006, p. 146). Thus, positivists see policy outputs as the outcome of a cause and effect relationship. Methodologically, positivists favour quantitative methods including econometric models, input–output analyses, and statistics (Armstrong and Wells 2006, p. 264). They strive to draw conclusions that can make generalizations possible (Mark et al. 2000, p. 162). Positivist approaches tend to be ‘top-down’, with the analysis focused on the macro-level; regional or local particularities and contexts are purposely ignored. Data is collected at an aggregate level, with individual firms or other participants generally reduced to passive actors or ‘inputs’ – a dialogue between stakeholders does not take place (Armstrong and Wells 2006, p. 264; Bachtler and Wren 2006, p. 146). Evaluators consider themselves perfectly neutral observers, independent from the process of policy implementation and detached from the parties involved (Mark et al. 2000, p. 144). A key characteristic in this respect is the assumption of a dichotomy between values and facts – absolute numbers are favoured. Such ‘black and white’ evaluations are frequently criticized for failing to measure the true impact of redistributive or investment programmes; other exogenous factors are ignored because they would blur the simple logic of cause and effect; the notion of causality, linking financial inputs to policy outcomes, provides the rationale for engaging in evaluation.

Constructivism
More recently, constructivist thinking has impacted upon evaluation studies. Armstrong and Wells (2006, p. 266) argue that ‘it is hard to see how the positive evaluation of the type prevalent in the SF can ever be compatible with a constructivist philosophy’. Thus, constructivism in evaluation research holds that facts and values cannot be clearly separated and that reality is always constructed, even though it may not be possible to assert how (European Commission 2008, p. 21). Objective knowledge is generally rejected; there are only subjective opinions and observations (Bachtler and Wren 2006, p. 146). In other words, the outcome of evaluation always depends on several factors specific to the particular situation; theory merely seeks to simplify the chaos of the external social ‘reality’ (Armstrong and Wells 2006, p. 266).

Methodologically, constructivist evaluators prefer case studies, following a so-called ‘hermeneutic-dialectical method’, which requires the assumptions and ‘constructions’ of stakeholders involved to be uncovered and confronted, and as such, it takes a bottom-up approach. The evaluator places particular emphasis on the local situation and its historical and cultural dimension (Armstrong and Wells 2006, p. 266). According to Guba and Lincoln (2001, p. 2) – strong proponents of the constructivist approach – the evaluator’s responsibility is to avoid automatically favouring those views held by the authority contracting the report (managerialism), since it may disenfranchise other stakeholders involved (evaluated) (Guba and Lincoln 2001, p. 8). This is vital to the epistemological core of constructivism – transactional subjectivism – for it means that ‘assertions about “reality” and “truth” depend solely on the meaning sets (information) and degree of sophistication available to the individuals engaged in forming these assertions’ (Guba and Lincoln 2001, p. 1). In short, evaluation design and practice are activities all stakeholders engage in, and contribute to; hence constructivists see evaluators themselves as subjects of the evaluation process (European Commission 2008, p. 21). The evaluator is expected
to defend his/her findings and adapt future assumptions according to what was learnt throughout the evaluation process.

**Realism**

Similar to constructivism, realism is an ‘umbrella’ under which many different theoretical approaches across disciplines can be found and can mean different things. While a US tradition sees it as a middle way between positivism and constructivism (e.g. Mark et al. 2000, p. 157), a ‘European’ school sees it as an alternative to dichotomies on methods, which tend to divide positivism and constructivism, seeing the two schools as incompatible. Pawson and Tilley (2004, p. 10) state that realist evaluation has ‘no preference for either qualitative or quantitative methods (...). The precise balance of methods to be used is selected in accordance with the realist hypothesis being tested, and with the available data’. Arguably, this approach positions realism between positivism and constructivism regarding methods and might be considered ‘somewhere in-between’, i.e. between positivists, who mostly use quantitative methods, and constructivists who only use qualitative methods. Thus, *epistemologically*, realism occupies a middle-ground between classical positivism and relativist constructivism.

Realists accept different accounts and categorizations in programme evaluation (Mark et al. 2000, p. 157). Their approach is generally concerned with finding out why things work in different contexts (European Commission 2008, p. 21). The evaluator is expected to generate theories and recognize patterns in the social and organizational processes that the programme has brought about (Armstrong and Wells 2006, p. 265). Methodologically, realism also holds an intermediate position, with the use of both quantitative and qualitative methods, depending on the particular situation. Evaluators use top-down and bottom-up approaches, depending on the context – for larger programmes a top-down macro-analysis might be more appropriate, whereas smaller interventions call for a bottom-up micro-study. Often combinations of the two approaches are chosen (Mark et al. 2000, p. 163; Pawson and Tilley 2004, p. 10). Realist approaches allow for the active involvement of stakeholders in the process, but evaluation is generally seen as a tool to serve the interests of programme managers (Bachtler and Wren 2006, p. 147). The object of study is an ‘open system’ whereby the evaluator should form a ‘teacher–learner relationship’ with programme officials and other participants (European Commission 2008, p. 21), hence the expectation that their concerns and suggestions feature strongly in the findings. Realist evaluators consider themselves to be part of the social reality but do not consider their role to impact upon their findings (Bachtler and Wren 2006, p. 147).

Table 2 provides an at-a-glance overview of the relevance or ‘fit’ between practical factors influencing practical approaches to, and theoretical perspectives upon, evaluation behaviour. It thus builds on table 1 which, identifies the general characteristics distinguishing different approaches to evaluation.

**CASE ANALYSIS: APPROACHES TO EVALUATING COHESION POLICY**

**Case selection and methodology**

Given the explorative and conceptual nature of the paper, small-N case studies seemed the most appropriate research design, appropriate for exploring or inferring a hypothesis, and on which initial findings can later be tested more rigorously with large-N studies (Van Evera 1997, p. 53). The choice of cases was made based on the notion of geographical scale. It sought to compare evaluations for a whole country (Ireland), region (North-Rhine Westphalia, Germany), and city or local region (Luckenwalde, Brandenburg,
TABLE 2 How the characteristics of practical approaches relate to theoretical perspectives

<table>
<thead>
<tr>
<th>Theory</th>
<th>Positivism</th>
<th>Realism</th>
<th>Constructivism</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Epistemological</strong></td>
<td>There is an objective reality</td>
<td>Several possible accounts, but not all equally valid</td>
<td>No objective ‘reality’ or ‘truth’</td>
</tr>
<tr>
<td><strong>assumptions</strong></td>
<td>Programmes either work or do not work</td>
<td>Facts and values should not be separated since both important for programme</td>
<td>Performance gauged by views, opinions, and experiences of stakeholders involved</td>
</tr>
<tr>
<td></td>
<td>Facts/statistics/quantifiable data only matters</td>
<td>Underlying structures governing workings of programme need to be deconstructed</td>
<td></td>
</tr>
<tr>
<td><strong>Methods/data</strong></td>
<td>Quantitative methods favoured, especially</td>
<td>Quantitative or qualitative (especially in-depth cases studies), depends on programme requirements and specific context</td>
<td>Case studies use the ‘hermeneutic-dialectical’ method, discussion with stakeholders in groups</td>
</tr>
<tr>
<td><strong>collection</strong></td>
<td>econometric models</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Top-down/bottom-up</strong></td>
<td>Top-down (macro-level, focus on generalization and leave particularities largely aside)</td>
<td>Top-down/bottom-up (macro/micro level)</td>
<td>Bottom-up (micro-level, take into account local/regional particularities, limited possibility to generalize)</td>
</tr>
<tr>
<td><strong>approach</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stakeholder</strong></td>
<td>Only passive as an economic actor</td>
<td>Evaluator and stakeholders form a teacher-learner relationship</td>
<td>Very high, even involving stakeholders reluctant to participate/unaware of their involvement</td>
</tr>
<tr>
<td><strong>involvement</strong></td>
<td>Object of analysis</td>
<td>Suggestions of the stakeholders form an essential part of evaluation report</td>
<td></td>
</tr>
<tr>
<td><strong>Role of the</strong></td>
<td>Does not get involved Takes a ‘God’s eye</td>
<td>The evaluator makes choices regarding categories and methods Deconstructs underlying mechanism but not actively involved</td>
<td>Evaluator strongly involved in project, makes his own constructions explicit</td>
</tr>
<tr>
<td><strong>evaluator</strong></td>
<td>perspective</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Germany). The rationale was that variations in the scope and scale of a practical evaluation exercise may reveal different research approaches – in short, we would compare different geographical scales and programme funding types.

Taking into account the particular characteristics of the different programmes, the following pattern might be expected: since the scale of funding relative to GDP is largest in Objective 1 areas, the focus of evaluation will lie on large scale, macroeconomic indicators, and the success of Objective 1 regions will be predominantly assessed using evaluation methods based on positivist evaluation theory (Bachtler and Wren 2006, p. 147; see also Esposti and Bussoletti 2008). In contrast, for Objective 2 regions, Structural Fund spending may be an important factor in stimulating economic change (Baslé 2006, p. 232), but the use of positivist or econometric methods is more difficult (Baslé 2006, p. 227).
Baslé suggests that in order to assess Objective 2 regions, explicit theories of the working mechanism of the Structural Funds in ‘Competitiveness and Employment regions’ should be created (Baslé 2006, p. 232), and implies the use of realist evaluation methods. Third, in the case of the URBAN II Community Initiative, a predominant focus is the ‘economic and social regeneration of troubled urban areas’ (Baslé 2006). Given these rather soft and broad socioeconomic goals, plus the local scale of the initiative, one might anticipate that constructivist evaluation methods, involving a high degree of interaction with citizens and a strong focus on subjective opinions, would be particularly appropriate. In this vein, Murtagh and McKay (2003, p. 195) stress the importance of the ‘discursive relationship’ between local people and managers of urban cooperation projects.

All Structural Fund programmes are to some extent goal-oriented, i.e. they seek to achieve certain objectives within an established timeframe – the programming period (of seven years). The most important question for the evaluator then, is how to set indicators to measure these goals? It is in this respect that strong differences between the different programmes occur: the goals of the Objective 1 programme are much more quantitative (usually recognizing an increase in GDP) than the goals of the URBAN II programme, which include the fight against social exclusion and the establishment of a stronger local identity (NIPP 2011); here goals are formulated in a ‘softer’ way and require indicators conducive to a constructivist approach.

We selected ‘most likely cases’ for each of the theoretical approaches. According to Gerring, ‘if a most likely case is found to be invalid, this represents strong disconfirming evidence’ (Gerring 2006, p. 213). As mentioned above, Objective 2 and URBAN II are most likely cases for realism and constructivism respectively, and simultaneously, least likely cases for positivism. Thus, under the present research design, the set-up was unfavourable for the assumption that positivism dominates. Hence, were we to find no application of realism and constructivism in the cases, we could argue generally they are seldom applied. In contrast, if positivism were dominant even in the least likely cases, then this would suggest positivism to be the dominant theoretical approach across the Cohesion Policy evaluation community in general.

The fact that two of the programmes were situated in Germany and one in Ireland is not deemed relevant, since in two cases the evaluations were carried out by multinational consortia consisting of public and private bodies. National evaluation culture might indeed still be a relevant factor but cross-cultural comparisons are not considered here. However, it has to be mentioned that no absolute generalizations can be made on the basis of three cases only. Instead, we set out to identify possible patterns in the application of theory.

**Objective 1 (national)**

**Epistemological assumptions**
In their methodological section, the evaluators stated: ‘analyzing a very large and complex programme of public investment, such as the National Development Plan/Structural Funds is not a simple task and there is no “correct” way to determine the appropriate prioritisation of measures’ (ESRI 2003, p. 83). Such a statement to some extent mirrors realist theory (Mark et al. 2000, p. 157), though might be considered as a ‘disclaimer’, since subsequent statements on the methods employed showed evidence of a more positivist stance. It was further stated: ‘in order to identify measures which are justified and those that are not, a screening method is particularly useful. It provides a means to formalize what
are otherwise purely subjective choices and ensures that all measures are treated equally’ (ESRI 2003, p. 88). The evaluators presumed to be making entirely objective statements.

Methods/data collection
Macroeconomic modelling and cost–benefit analysis were the only means of analysis, the HERMES model being the main tool for analyzing the impact of Structural Aid on GDP growth (ESRI 2003, pp. 52–63). The econometric analysis placed a strong focus on the quantifiable effects of the Structural Funds, though with some consideration to ‘horizontal issues’, such as social inclusion – the evaluators stated that ‘consultation with key informants and stakeholders’ was conducted in order to assess the effectiveness of the funds in this respect (ESRI 2003, p. 197). Qualitative interviews were used as an additional method but the report criticized their use because of the low quality of relevant indicators used when evaluating programme objectives (ESRI 2003, p. 194).

Top-down/bottom-up approach
The evaluation clearly took a top-down, meta-level approach and did not delve into particular local features. The approach was to conduct a macro-analysis of the whole economy, with data being collected on an aggregate level, hence a ‘removed’ positivist standpoint.

Stakeholder involvement
The report employed a rather narrow definition of ‘stakeholder’, the main part of the analysis treating companies, investors, consumers, and workers as mere ‘factors’ in the econometric models and statistics – and passive factors at that. Stakeholders’ opinions were not made explicit or published in the report, thus limiting the degree of reflexivity, and arguably, real feedback based on the human experiences of those engaged in programme implementation.

Role of the evaluator
Though a first-person position was sometimes taken, as indicated by the use of ‘we’ or ‘the evaluators’ (ESRI 2003, p. 166), in general no further clarification of the evaluators’ role was given. The philosophical approach taken by the evaluators predominately was that of offering a ‘God’s eye view’, as Putnam calls it (see Mark et al. 2000, p. 144), whereby the researchers are independent from the programmes and stakeholders, and hence objective.

Objective 2 (regional)
Epistemological assumptions
The evaluators made their epistemological assumptions explicit by stating: ‘Establishing the efficiency of a programme is always a subjective appraisal’ (EPRC et al. 2003, p. 1). This is in line with the realist conception of evaluation research, as a broad endeavour aimed at discovering the different mechanisms and conditions that make a programme work; and thereafter, about generating hypotheses pertaining to why a programme succeeded or failed. In such ‘theory driven evaluation’, socio-cultural aspects were taken into account and recognized as preconditions for a climate of innovation, thus reflecting the doctrine of ‘alternative accounts’ found in realist evaluation theory (Mark et al. 2000, p. 257).
Methods/data collected
Typical quantitative-positivist methods like SWOT (Strength, Weaknesses, Opportunities, Threats) – a strategic planning approach traditionally used in business consultancy – were used (EPRC et al. 2003, p. 41), alongside statistics on unemployment and quantified data regarding sectoral developments over time (EPRC et al. 2003, p. 45). Nevertheless, these techniques were accompanied by a qualitative approach, with 70 semi-structured interviews and 11 in-depth case studies conducted, albeit only among elite groups directly involved in the programme.

Top-down/bottom-up approach
Evaluators gauged the opinion of a constellation of actors, to understand the particularities of the cases. They took into account socio-cultural aspects, which they perceived as being highly relevant to the success of a programme (EPRC et al. 2003, p. 18). The report emphasized the results of interviews and case studies, framed them as key to improving the programmes (EPRC et al. 2003, p. 16). Thus, much emphasis was placed on micro-level analysis and on data collected directly from stakeholders. Evaluation took a bottom-up, realist perspective.

Stakeholder involvement
The many interviews conducted at various levels point to the broad involvement of stakeholders, but did not necessarily indicate that anything valuable was gleaned from the exercise. Interviewees included civil servants at the ministerial public administration in the Land responsible for implementation, at the office of the region’s ‘prime minister’, and other ministries (EPRC et al. 2003, p. 441). The local level was involved through interviews with representatives from cities in the region (EPRC et al. 2003, p. 442). Firms and institutions that benefitted directly from the programme (i.e. were recipients of ERDF) were consulted, along with several academics. However, target groups such as NGOs or the social partners who were not contributing financially, or merely acting as observers, were not consulted, i.e. stakeholder involvement followed a classical positivist paradigm, even though a realist position was sought. Yet, the evaluators regretted a general lack of reflexive involvement in the evaluation process – ‘evaluation is still as a form of control and not as a learning tool’ (EPRC et al. 2003, p. 17).

Role of the evaluator
Despite the neutral tone adopted throughout the report, emphasis was placed on the importance of the working relationship between evaluator and stakeholders: ‘that the “setting” as a whole is seen as highly important for facilitating learning processes, presupposes that decision-makers (persons and institutions) allow themselves and the evaluators to assess their weaknesses, and are willing to change; evaluation thus requires a strong climate of trust’ (EPRC et al. 2003, p. 19). The validity of the evaluation process and worthiness of the findings depended on conveying empathy and forging a strong human bond. The evaluators saw themselves as integral to the evaluation process, influencing it through their behaviour, i.e. they were part of the social ‘reality’ (Mark et al. 2000, p. 152). In short, claims of programme success made by the evaluators rested on their own subjectivity and how they perceived the programme’s performance (EPRC et al. 2003, p. 19).
Community initiative (local)

Epistemological assumptions
The evaluation report comprised two parts – a main report and an additional sociological analysis on urban identity. While the former had a positivist outlook, with the aims defined as ‘effectiveness, efficiency and effect on horizontal goals’ (Isoplan Marktforschung 2003, p. 6), the latter’s aim was to examine ‘the extent to which citizens identify with their city’ (Stadtbüro Hunger 2003, p. 4), by analyzing the ‘perceptions, views, interest and ideas of the citizens’ (Stadtbüro Hunger 2003, p. 4). Hence, the evaluation adopted a very reflexive approach, attributing importance to subjective opinion and even feelings, closely in line with a constructivist approach.

Methods/data collected
Quantitative methods were used to calculate the money spent implementing projects, and how the local labour market had been evolving. By contrast, the sociological report employed only qualitative methods: elite interviews with a small number of stakeholders, followed by a representative survey among households, and complemented with a special survey among local schoolchildren (Stadtbüro Hunger 2003, p. 7). The use of interviews and surveys suggests a realist approach. Though importance was attributed to the values and perceptions of citizens (Stadtbüro Hunger 2003, p. 4) they did not ‘feed-back’ into the design of the evaluations. A genuine constructivist approach would require the initial design (or ‘construction’) of the evaluation project to be revised on the basis of stakeholders’ feedback, as well as their preferences and expectations of the project, as Guba and Lincoln (2001, p. 5) would argue.

Top-down/bottom-up approach
For the main evaluation, elite interviews mostly concerned those directly involved in the programme at upper levels, for example the public administration responsible for the Structural Funds at the Ministry of Finance in the German Land, the body implementing URBAN II, plus city administration staff (Isoplan Marktforschung 2003, p. 14). This narrow selection of interviewees mirrors a positivist approach. However, for the identity study an array of citizens was consulted, including local politicians, social workers, representatives of industry and chambers of trades and commerce (Stadtbüro Hunger 2003, p. 4) – evidence of a targeted (constructivist) approach at the lower levels. The large survey (N = 449) captured the views of ‘ordinary citizens’, even children (Stadtbüro Hunger 2003, p. 5).

Stakeholder involvement
The extensive consulting and surveying of a large sample group suggests a bottom-up approach. The focus was on the micro-level, with evaluators seeking to gauge feeling, ideas and perceptions (positive or negative) about the future at street-level (Stadtbüro Hunger 2003, p. 4). In the main evaluation report, both bottom-up methods (interviews, personal visits to the site) as well as top-down methods (desk-based research of aggregate data) were used (Isoplan Marktforschung 2003, p. 15).

Role of the evaluator
The role of the evaluator differed from the mainstream Objectives 1 and 2 programmes. For the main report the evaluators visited the town and project sites, ‘in order to gain an impression of the preconditions and the state of implementation’ (Isoplan
### TABLE 3  Summary of the comparative analysis of Structural Funds programme evaluations

<table>
<thead>
<tr>
<th>Programme/Indicator</th>
<th>Objective 1 (Ireland)</th>
<th>Objective 2 (NRW)</th>
<th>URBAN II (Luckenwalde)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epistemological assumption</td>
<td>Positivism</td>
<td>Realism</td>
<td>Positivism/Constructivism</td>
</tr>
<tr>
<td></td>
<td>Strong focus on effectiveness, no acknowledgement of alternative accounts</td>
<td>Acknowledgment of subjectivity of every evaluation and of alternative accounts, wide definition of effectiveness</td>
<td>Difference between identity report and main evaluation: focus on effectiveness and stakeholder perceptions respectively</td>
</tr>
<tr>
<td>Methods/data collection</td>
<td>Positivism</td>
<td>Positivism</td>
<td>Positivism/Realism</td>
</tr>
<tr>
<td></td>
<td>Econometric models (HERMES), statistics, SWOT analysis, elite interviews</td>
<td>Econometrics and statistics; elite interviews</td>
<td>SWOT analysis, elite interviews; in identity report: elite interviews and surveys (households and students)</td>
</tr>
<tr>
<td>Top-down/bottom-up approach</td>
<td>Positivism</td>
<td>Realism</td>
<td>Realism/Constructivism</td>
</tr>
<tr>
<td></td>
<td>Macro perspective, regional and local circumstances not taken into account</td>
<td>Micro perspective, the constellation of actors and particularities of the cases is taken into account</td>
<td>Main evaluation: top-down (macro perspective); identity report: bottom-up, focus on local particularities and perceptions of the citizens (micro-perspective)</td>
</tr>
<tr>
<td>Stakeholder involvement</td>
<td>Positivism</td>
<td>Positivism</td>
<td>Positivism/Constructivism</td>
</tr>
<tr>
<td></td>
<td>Narrow definition of stakeholders in main analysis, merely objects for econometric analysis, wider definition of stakeholders in evaluation of horizontal issues</td>
<td>Relatively broad definition of stakeholders, but no involvement of social actors, teacher–learner relationship between evaluators and stakeholders aimed at</td>
<td>Persons directly involved in the programme; in identity report: wide definition of stakeholder and ordinary citizens (survey); feelings, ideas and perceptions of the citizens are taken into account</td>
</tr>
<tr>
<td>Role of the evaluator</td>
<td>Positivism</td>
<td>Realism</td>
<td>Realism</td>
</tr>
<tr>
<td></td>
<td>Use of first person (‘we’), but no explicit acknowledgement of role of evaluator</td>
<td>Neutral tone, setting of the evaluation seen as important, relationship of trust between evaluator and stakeholders</td>
<td>Visit of evaluators at the site and strong emphasis of perceptions of citizens, no reflectivity, role of the evaluators not made explicit</td>
</tr>
</tbody>
</table>

Marktforschung 2003, p. 15). This high degree of personal involvement and engagement with the ‘fieldwork’ indicates a realist approach, as supported by the additional study analyzing participants’ values and aspirations (StadtBüro Hunger 2003, p. 3). Evaluating
a Community Initiative for a small population saw the mixed use of realist and positivist approaches by an ‘involved’ evaluator making the most of face-to-face interactions.

Table 3 shows how the different theoretical approaches (as illustrated in table 2) are mirrored in the evaluation reports of the three cases.

Interim conclusion
As the analysis has shown, elements of constructivism and realism could be found in the evaluations of the Objective 2 and URBAN II programmes. However, positivism is on balance the dominant approach in all three cases. Objective 2 and Urban II were considered most likely cases for realism and constructivism, respectively, and least likely cases for positivism. We can confirm that positivism is indeed dominant in the evaluation of Structural Funds and that the other approaches are rarely applied. While this may not come as a complete surprise, it does reveal a stark contrast with the rhetoric of the European Commission, which advocates a combination of different evaluation approaches (European Commission 2008, p. 97). The paper for the first time empirically confirms, be it through a small but representative case study, the widely held ‘gut feeling’ in evaluation circles that positivism is dominant. In the next section we discuss possible reasons why this is so, in line with the five characteristics of approaches to evaluation identified earlier in the paper.

EXPLAINING THE DOMINANCE OF POSITIVISM IN EU POLICY EVALUATION

Epistemological assumptions
Positivism fits the epistemological assumptions widespread in the Commission concerning what represents proper knowledge and sound data. Given the huge value of the Structural Funds it is essential to be able to demonstrate that programmes achieved (or not) what they set out to achieve (Molle 2008, p. 224). For Cohesion Policy implementation to be poorly perceived would ‘constitute a threat for the Commission’ (Bauer 2006a, p. 721). Gore and Wells (2009, p. 159) indicate how the Commission sees the world: ‘what is striking about the guidance for the mid-term evaluations is its tight focus on issues of management processes, programme relevance and the narrow assessment of programme objectives’. From a principal–agent perspective, in order to compensate an inferior informational position vis-à-vis its implementing agents (local and regional public administrations), the Commission invests in strict monitoring mechanisms (Bauer 2006a, p. 722). Quantitative data is suitable and convincing, allowing for easy aggregations, comparisons, and generalizations.

Stakeholder involvement
The Commission, like the EP and the Court of Auditors, favours quantitative data because of its usefulness in communicating the added-value of EU policies. The institutions must also demonstrate the ‘additionality’ of funds, i.e. that programmes/projects could not have been carried out with national monies alone, nor the results achieved without EU budgetary contributions (Molle 2008, p. 150.). For political reasons it is important to demonstrate a high absorption rate, i.e. that most of the Structural Funds available to regional authorities was spent on projects and not ‘de-committed’ (Molle 2008, p. 196), to be returned to the Commission’s ‘common pot’. The emphasis is on demonstrating the maximum engagement of stakeholders active in EU policy at street-level as eligible recipients of EU funds; the nature or quality of participation is arguably less important.
than the quantity. Figures can be used to compare stakeholder involvement (i.e. how much money ‘flowed back’ to them as project partners), whereas words are opaque, can be interpreted differently, and create ‘shades of grey’.

**Costs of methods and data collection**

Methods advocated by realist and constructivist theory, such as surveys, interviews, or meetings with stakeholders in small groups, are more expensive to carry out than positivist methods such as economic modelling – basically a desk-based activity seeking out pre-existing data. Evaluation is paid from the Technical Assistance budget line of a programme budget. It was recalled in the mid-term evaluation guidelines that ‘any disproportionate expenditure should be avoided’ (European Commission 2000, p. 10). Financial restraint in evaluation is welcomed by the member states, who prefer money spent on investment projects for tangible results, some considering evaluation a ‘necessary evil’ and a performance of ‘going through the motions’.

**Top-down/bottom-up approach**

It is arguably easier to carry out evaluation in a top-down/centralized manner. Some regional officials believe that the Commission sees evaluation as a ‘supranational instrument for controlling policy implementation’ and exerting pressure on regional public administrations (Bauer 2006b, p. 781). Numbers and percentages can be compared from one programming period to the next, and across geographical areas. It is thus arguably much easier to compare the impact of cohesion policy across countries using positivist methods than using other approaches, without criteria and indicators varying between countries. ‘Hard’ numbers also make it easier to decide whether a programme should receive additional funding from the performance reserve. Reports conducted using realist or constructivist techniques are more difficult to use when seeking to sanction implementers for not complying with set (quantified) targets, goals, and deadlines. Nonetheless, national and regional administrators do not necessarily share the enthusiasm for quantitative evaluation, often doubting the scientific reliability of results and lamenting the bureaucracy involved in ‘number fetishism’ (Bauer 2006b, p. 109).

**Role of the evaluator/the evaluation community**

Finally, the dominance of positivism may be intrinsically linked to how the evaluation community sees its own role. Many influential scholars in the field adhere to a ‘rationalist’ or positivist idea of evaluation (Howlett and Ramesh 2003, p. 219). Some authors claim the first conceptualizations of evaluation adopted a positivist outlook (Campbell and Stanley 1971), hence an element of path dependency – tried-and-tested methods tend to ‘stick around’. Many senior evaluation scholars were educated at a time when positivism was more or less unquestioned as a social science paradigm (Campbell and Stanley 1971, p. 241). As a consequence, most leading evaluation firms, and companies offering evaluation services such as accountancies and consultancies, generally favour quantitative/econometric approaches. The preferences of institutions tendering out evaluations (Commission and national ministries) also often match those of the consortia bidding to win evaluation contacts – if preferences do not match, no deals would be struck. EU evaluation is big business conducted in a political subsystem of political, financial – and even reputational – inter-dependencies.

In short, should we be concerned by the dominance of positivist approaches to EU policy evaluation? Stame (2008, p. 128) raises an important point, arguing that such a
narrowly-focused approach may deprive evaluators of opportunities for triangulation. Mark et al. (2000, p. 163) also acknowledge the importance of combining different theoretical approaches to improve evaluation quality. Bauer (2006a, p. 729) argues that an exclusive focus on quantitative results and indicators may actually result in decreasing information flows, i.e. we end up knowing less not more about what really happened.

The negative consequence of an automatic and self-reinforcing focus on positivist method can lead not only to unsatisfactory evaluation outcomes, but even to the failure of the programme itself (Kelly and Maynard-Moody 1993, p. 141). A stronger emphasis should arguably be placed on the various post-positivist approaches to policy evaluation, such as realism and constructivism, in order to make evaluation more effective and legitimate (Fischer 1998, p. 140).

CONCLUSION
The article has shown up marked characteristics in the approach to Structural Funds evaluation. The analysis has highlighted the dominance of positivist approaches to evaluating Cohesion Policy and sought to account for it. By developing a toolkit that combines real-life characteristics pertaining to the task of policy evaluation, with theoretical perspectives from political science, it has provided insights into how theory might be better linked to practice for EU policy analysis. More reflective thinking about evaluation would have valuable implications at policy level. A more pronounced use of mixed approaches would arguably strengthen the effectiveness of evaluation in terms of assessing how policy fared at programme level and the nature of its results and impacts – which are often intangible or identity-centred and thus difficult to evaluate using positivist approaches. However, given the political and ideological factors favouring the use of positivism by key political, institutional, and private sector interests engaged in policy evaluation, a widespread application of such approaches seems rather unlikely.

The evident misfit between programme implementation and the approach to evaluation raises many questions for future research, such as: How to make better use of qualitative data, even in positivist evaluation design? How to better demonstrate effectiveness? How to ensure the comparability of data collected by different methods? Such questions could fruitfully be examined in public policy research that pays more explicit attention to the interface between theory and practice, while acknowledging how the choice of evaluation approach and design from the very outset determines and constrains findings, with implications for policy reformulation, the EU budget and the political fate of institutional actors across the EU’s multi-level administrative system.

ACKNOWLEDGEMENT
The authors would like to thank four anonymous referees for their critical and highly valuable comments.

REFERENCES

Bauer, M. 2006b. ‘The EU “Partnership Principle”: Still a Sustainable Governance Device Across Multiple Administrative Arenas?’ *Public Administration*, 80, 4, 769–89.


**Date received 3 February 2011. Date accepted 6 October 2011.**