Theory Based Evaluation: A wealth of approaches and an untapped potential

Marielle Riché
European Commission[i]

"Have we got closer to identifying the generative mechanisms that produce an outcome for the lagging regions of the EU? The answer to this last question in the opinion of the author is a resounding no. Too often the cry from the evaluation community is for more accurate data (usually quantitative) upon which to apply existing methodologies. Than misses a fundamental point that there is a real need to embrace more appropriate methodologies which are just as data demanding but in a rather different way". Mark Hart, 2007

Abstract

Since the 1990s, many works have refined the Theory based approach dressing it up with new names such as Contribution Analysis, Elicitation Method or General Elimination Methodology. Theory Based Evaluations are particularly relevant for policy makers as they explain why an intervention works -or not- in a given context allowing for generalisation. Their potential is not sufficiently exploited under the programmes supported by the European Regional Development Fund (ERDF) for two main reasons: programmes rarely articulate a clear theory of change. Their evaluations do not build on rigorous theory-based frameworks integrating and guiding the use of a mix of methods.

KEYWORDS: programme theory, context, policy replication, systematic reviews, impact evaluations, regional policy, quality standards and criteria

Introduction

Since the beginning of the programming period for 2007-2015, some 266 evaluations have been carried out throughout Europe concerning programmes co-funded by European regional policy. These programmes, which are 9 years long, support socio-economic and regional development and fund varied interventions ranging from environmental, health, telecommunication or transport infrastructures, urban generation and tourism services, to financial support or services to companies, research or networking activities. They may operate regionally or nationally, and their evaluations may cover a geographic area, a sector of activity, a group of persons as well as a major project or a portfolio of interventions.

Within the European Commission and its directorate general "Regional policy", an Evaluation unit is dedicated to commissioning *ex post* evaluations after each programming cycle as well as thematic evaluations where deemed necessary for EU policy making. Other tasks are to support programme Managing Authorities in raising the quality of their evaluations, to collaborate with other international organisations and with academics working in the field, and to encourage the exchange of practices amongst European evaluation practitioners.

Developing evaluation good practice throughout Europe

From the 1990s onward, EU regulations imposed an obligation to undertake evaluations of regional policy programmes. This requirement, combined with broader trends in policy making such as "new public management" and "evidence based policy" which spread from the UK and the USA to most of the European countries during the last twenty years, contributed to develop an evaluation culture in Europe. Except from some countries such as the UK, Germany and Sweden, most Member-States were not evaluating their regional policies on a regular basis. EU regulations not only improved managerial practices in ERDF programmes with the implementation of systematic monitoring and evaluation, but also induced spillovers on domestic management practices[1]. This also helped to develop an evaluation market in EU Member-States, still underway in countries having recently joined the EU.

Judging the design and findings of regional policy evaluations is not an easy task, as the terms of references and details on methods and budgets are rarely published. However, we have had several opportunities to quality assess some, for example in 2004 when reviewing all the programmes' mid-term evaluations to attribute the so-called "performance reserve" to Member States[2]; through the work of an expert evaluation network set up in 2010 to report each year on Regional policy performance including evaluations[3]; and in the context of case studies of evaluations covering publicly supported innovation activities conducted for us in 2011[4]. These reviews show that the overall quality and use of these evaluations has steadily increased, albeit with significant variations between Member States.

In particular, it is clear that few evaluations focus on the outcomes of the programmes (27%); most are concerned with management and implementation issues. Concerning the design, most evaluations are based on qualitative techniques and on monitoring indicators. Furthermore, the same methods seem to be favoured at the expense of more recently developed techniques that could better respond to the specific evaluation questions, e. g. social network analysis in the field of evaluations of innovation activities or experimental approaches to assess the impact of ERDF interventions.

Since 2008 we have carried out a number of activities to develop impact evaluations and the experimentation of new methods in our programmes. We did not start in isolation as in the same year, the World Bank was mainstreaming its "Development

[1] See ex post evaluation of the ERDF, workpackage 1: Coordination, Analysis and Synthesis http://ec.europa.eu/regional_policy/sources/docgener/evaluation/expost2006/wp11_en.htm

[2] One criteria for allocating this performance reserve was the quality of mid-term evaluations

[3] See report 2011 on http://ec.europa.eu/regional_policy/information/evaluations/index_en.cfm#1

[4] See Technopolis Group & MIOIR study, 2012

Impact Evaluation Initiative" and in 2009 the network NONIE had published the guidance of impact evaluation[5]. We first worked in collaboration with experts in the field[6] to understand the relevance and limit of experimental techniques for the kind of activities our Funds support. In 2010 and 2011, we launched four evaluations to test comparison group techniques on grant schemes to companies in different settings. We also supported the organisation of two summer schools to train Managing Authorities willing to experiment with these techniques in their evaluations the same years. A note summarising the results of these evaluations and their implications for policy will be published soon on the Directorate General's website[7]. In parallel, we developed our expectations for the use of case studies in the framework of ex post evaluations 2000-2006 which all integrated regional case studies in their design. We have also experimented with the use of ex post cost benefit analysis for large scale infrastructure projects. While recognising that non-experimental methods[8] using control or comparison groups, which assess whether particular interventions work or not, are very useful tools for appraising cost effectiveness and questioning policies, we have understood that their applicability to ERDF interventions was limited (need for well specified and large scale homogeneous interventions) and that they were not sufficient *per se* for policy learning.

Most of our programmes encompass activities of various natures, often interlinked and contributing to different development objectives. Some focus on particular territories or groups of people, meeting their needs through "integrated" approaches that address simultaneously different aspects of a problem, cover various sectors of activity and involve different actors. Furthermore, taking place in different settings, our programmes and their outcomes are influenced by a diversity of social, institutional, cultural, economic, geographical and political factors making it crucial to understand the "why it works" beyond the "does it work". Besides knowing if an intervention works, policy makers need to understand how to improve or replicate it, i.e. they need contextualised evaluations.

This is also a prerequisite for building policy knowledge across Europe. In this respect, our long-range ambition is to develop a still neglected dimension of regional policy evaluation activities, which is capitalising on their findings and building shared knowledge on similar underlying mechanisms leading to some kind of regularities of outcomes in different contexts (the "middle-range theories" theorised in sociological literature). Such a knowledge base would also describe existing alternatives working in different contexts, favouring policy adaptation when policy replication is not possible.

[5] http://www.worldbank.org/ieg/nonie/guidance.html.

[6] Such as Alberto Martini, of Piemonte University, Robert Walker of Oxford University, Arianna Legovini of the World Bank, Dirk Czarnitzki of Leuven University.

[7] And a paper based on these findings is also being presented by my colleague, Daniel Mouque, at another session in this conference.

[8] "Non-experimental methods" which use statistical techniques are considered less rigorous than experimental methods based on "Randomized Control Trials". They are also labelled "quasi experimental".

Up to now, Managing Authorities of ERDF programmes cannot benefit from policy evidence such as those gathered by "The Cochrane Collaboration" working within the health field or "The Campbell Collaboration" in the field of education, crime and social welfare. This becomes more important with the stronger requirement of evidence-based policies. However, to be able to build similar repositories on the different fields of activity funded by the ERDF programmes, we need evaluations not only producing strong evidence on individual programmes' outcomes but also explaining the underlying mechanisms and how the context may influence these outcomes. This calls for approaches mixing experimental and qualitative techniques, which theory-based evaluations can encapsulate.

To develop the awareness and knowledge of what theory-based evaluations can offer, we recently asked Frans L. Leeuw, from the Maastricht University, to draft guidance outlining under which circumstances and how this approach can be applied (Leeuw, 2012). We also asked Frank Vanclay of University of Groningen to explore for us the possible use of various qualitative evaluation methods in our programme evaluations, in particular Most Significant Change Technique, Performance Story Reporting or Collaborative Outcomes Reporting (Vanclay, 2012). These studies outlined the potential for theory-based approaches to foster policy making in the context of regional policy programmes.

Theory-based evaluations are relevant for regional policy

Regional development programmes co-funded by the ERDF are usually bottom-up, aiming to meet specific needs of territories, sectors or group of beneficiaries in order to foster regional growth and employment. These programmes are based on the socio-economic analysis of local needs and potentials while being framed by higher goals set up in European Union strategies. When designing them, Managing Authorities are asked to explain *ex ante* their theory of change: how the expected change will contribute to both the local and European objectives (programme theory) and how the planned activities will lead to the expected results (implementation theory) [9].

During a programme's life, theory based evaluations are useful to assess if it progresses as planned towards outputs, intermediate and end outcomes, how it does it, and if it takes into account limiting factors in its implementation theory. They may analyse the results of experimental interventions launched to test new approaches. They also may examine the probability of programme effectiveness by analysing "whether the outputs include features that are preconditions to the achievement of the goals according to the intervention theory" ("Recently Implemented Policy Instrument" approaches described in Leeuw, 2012).

[9] See Weiss C. H., 1997:" Implementation theory focuses on how the programme is carried out to deliver the desired results[…] Programmatic theory, on the other hand, deals with the *mechanisms* that intervene between the delivery of program service and the occurrence of outcomes of interest. […] The mechanism of change is not the program activities per se but the response that the activities generate".

Ex post, these evaluations can validate the assumptions underlying the programme theory against the actual outcomes by getting inside the "black box" and analysing the mechanisms leading to these outcomes (including unintended effects) and the specific context in which they work. Here comes the debate between the advocates of Randomised Control Trials who claim that only experimental methods can rigorously assess impacts and proponents of Theory Based approaches that insist on the partial usefulness of experimental (or non-experimental) methods in socio-economic evaluations. In our opinion, this debate is missing the point.

First, non-experimental techniques are better adapted to operations involving a sufficient number of comparable entities to draw large enough samples. They are not applicable to important fields of intervention of ERDF programmes such as transport infrastructures (difficult to establish the counterfactual), innovation (interventions are not uniform, outcomes are uncertain, difficult to trace and long term) or complex programmes mobilising different simultaneous operations to address territorial or urban needs. In particular in richer Member States where Managing Authorities are asked to concentrate their limited ERDF resources on activities supporting research and innovation, many programmes aim to modify actors' behaviours and to strengthen innovation systems. Experimental methods can be used to establish causal links and estimate the impact of specific interventions where the preconditions are met, including the availability of data to build samples. However they will not be sufficient to explain the mechanisms influencing behavioural change in the desired direction or to assess the impact of complex interventions where outcomes may not be predictable and where experimentations are part of policy making.

Second, as already stated, policy makers need policy relevant evaluations, i.e. not only assessing the "does it work" but also capturing the "why" and "how". Understanding how the programme's interventions combine within the specific context, how they interact with other public interventions, programmes or policies, is of outmost importance to optimise the use of public money. As stated in a recent study commissioned by the UK Department of International Development, "*counterfactuals answer contingent, setting-specific causal questions 'did it work there and then' and cannot be used for generalization to other settings and timeframes, unless they are accompanied by more fine-grained knowledge on the causal mechanisms actually operating within the process leading from potential cause to effect*" (Stern et al., 2012).

Third, if policy makers request strong evidence, they also need narratives explaining how the programme achieved its results. This narrative supports policy understanding and communication to stakeholders, it is often required to convince policy makers about the rationale for modifying or replicating an intervention.

A major strength of theory-based evaluations for policy making is their formative and learning dimension. In some countries such as Sweden, this approach fits with a recent emphasis on on-going evaluation as learning tools for Cohesion programmes. This approach is "*based on dialogue, interaction and participatory inquiry as central elements*" (Brulin, Svensson, 2012). Through the joint knowledge formation between stakeholders, decision makers and evaluators, these evaluations will also contribute to the long term sustainability of programmes' outcomes.

More generally, regional programmes supported by the ERDF represent a major part of EU support to Member States. They are managed locally in cooperation with the European Commission. Through this "shared management", new management ideas such as local actors' involvement ("partnership principle") progressively modify national and local governance. Theory based evaluations, which make policy assumptions explicit, clarify policy goals and deploy techniques involving stakeholders – interviews , story-telling , participatory approaches, etc. – can contribute to change perceptions and behaviours from the design right through to the end of a programme, thus influencing its outcomes and ultimately policy thinking and practice.

An untapped potential

In spite of this potential for policy development, theory-based approaches are not sufficiently used in ERDF programmes. Several factors can explain this situation: the institutional/regulatory context to begin with.

ERDF programme implementation (and all Structural Funds) is framed by regulatory requirements that pressurise Managing Authorities into spending funds two years after committing them. If they fail, the funds come back to the EU budget. Therefore many evaluations are launched early in the programming period to analyse the programme implementation process with the aim to insure full and regular absorption of EU funding. Some also look at the continuing relevance of the programme objectives against a changing context. We have not found any example of ERDF evaluations looking thoroughly at the theory of change at this stage of implementation.

Until 2007, the Structural Funds Regulation imposed mid-term evaluations conceived as important milestones in the programmes' life. One drawback was that for some programmes facing delays in their implementation, these mid-term evaluations were undertaken too early to be useful. Another was their broad coverage (the entire programmes), encouraging general evaluation questions that, within the assigned budget, no evaluator could sufficiently deepen. A third point was that being required by the regulation, these costly evaluations were often the only ones carried out by the Managing Authorities for the whole period. Furthermore, up to now, the Regulation has not make it compulsory to demonstrate the impact of the programmes[10], leading to a concentration of evaluations during the first half of the programming period and very few focussing on outcomes, the latest evaluations being used to prepare the next round of programmes.

[10] Ex post evaluation is the responsibility of the European Commission and not obligatory at programme level. Only very few programmes have been subject to ex post evaluations by national or regional authorities. With the overlap of programming periods also, ex post evaluations of the previous programme period tend to be available only mid way through the next.

Another factor, probably the most detrimental to the development of theory-based evaluations, is that most programmes are not built on a clear intervention logic. They express broad socio-economic needs and objectives, but do not clearly specify what change is expected, how they will achieve it and in what timeframe. This situation may result from an intention to build sufficient flexibility into a programme in order to avoid the administrative work involved in a formal modification. It may also result from a natural tendency to consider that what was done in the past remains valid, without any attempt to question it: "evidence based policy" is still political rhetoric rather than a reality in many Member States.

Furthermore, Managing Authorities may consider that the overriding European or National priorities justify investment in some activities without the necessity for further evidence (innovation support is a typical field of intervention where it is taken for granted that, for example, financial support to SMEs and Research institutions will ultimately lead to innovation outcomes). Such an approach favours implementation evaluations (does the programme deliver as planned and how to improve its effectiveness?) to the detriment of evaluations testing the programme theory of change (do the interventions bring the expected benefit for the region or for the targeted sector or group of people, and how?).

When Managing Authorities design the result indicators required for each programme by the ERDF regulation, these indicators rarely reflect the policy objectives. Most actually measure the first results of the interventions (directly linked to the outputs) disconnecting them from the expected policy change. For example, a French regional programme aiming to increase the competitiveness of micro and small enterprises considered the number of coaching visits to SMEs "thus put in the path toward innovation" (*sic*) to be the outcome of the intervention. Another issue is the over-confidence put in indicators to measure the actual outcomes of a programme which marginalises the role of evaluation. Consequently many Managing Authorities do not take into account evaluation data needs in their programme design.

This situation may also reflect reluctance from policy makers to commit to socio-economic outcomes given the tiny share of structural funds in the overall regional development investments and the probability that they may make no measurable difference. Some Managing Authorities underline the difficulty to control the outcomes of activities often delegated to intermediaries such as development agencies or venture capital funds, others invoke the technical difficulty to follow the effects of the programme on broader group of entities than the direct beneficiaries because of insufficiently developed statistical systems at regional level and the cost of systematic surveys to fill this gap.

2014-2020: a new opening for more theory-based evaluations?

We believe that the next programming period 2014-2020 will create better conditions for theory-based approaches. This new period is characterised by a political commitment to design a "result-based policy"[11] for growth and competitiveness. During the last twelve months, we have prepared the ground for a step improvement in the design of the future programmes. We presented a new concept of the logical framework to our partners in the Member States and elaborated on it in a guide posted on our website since November 2011[12]. The proposed regulation on Cohesion policy programmes for the period 2014-2020 adopted by the European Commission includes an expanded list of points that the *ex ante* evaluators should analyse, including the theories of the draft programmes[13]. Involving the evaluators early enough in the design of the programmes should considerably improve their quality. They should help to elicit and make explicit programmes theories, constructively criticise common sense assumptions such as "training will improve employability", or "grants will increase market innovation", sometimes challenge them by proposing alternative theories based on past evaluations or research. They should also help to make theories testable and programme amenable to evaluation. We drafted a guidance document on ex ante evaluations to inspire the call for tenders due to be launched during third quarter 2012[14].

In parallel, we worked with some 14 volunteering Managing Authorities to test in a practical way the implementation of result-based approaches in their current programme, i.e. asking them to review their current programme theory and accordingly the result indicators reflecting the expected outcomes. To be relevant, these indicators need to be embedded into a clear and well-articulated programme theory. One difficulty we faced was to make clear the difference between "implementation theory" and " programme theory". Most Managing Authorities focus on implementation, committing to output or results directly linked to the interventions. As explains Rogers, this "*can be a good start, particularly for organisations unaccustomed to focusing on outcomes rather than activities*". However, this is not sufficient in a context of limitation of public funds where policies, including EU cohesion policy, are scrutinised to determine whether they add socio-economic value to European development. Our position is to push for programmes which state what they want to change in a territory, for an economy or a group of people and articulate the theory leading to this change.

Since 2007, ERDF programmes benefit from more flexibility to develop theory-based evaluations. Except for the *ex ante* evaluations, the timing and type of evaluation is not imposed to Member States by the Regulation. After the disappointing experience of the compulsory Mid-term evaluations in 2003, we preferred to leave up to the

[11] See the draft regulations for the 2014-2020 period adopted by the European Commission in October 2011 and the 5th cohesion report adopted by the Commission in November 2010

[12] European Commission Guidance document, 2011

[13] Article 48

[14] http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/ex_ante_en.pdf

Managing Authorities to decide on what kind of evaluations they needed and when. This approach was correct: if some Managing Authorities limited the number of their evaluations as compared to the previous period, most used them to support their programme implementation, review the continuing relevance of the programme interventions or deepen the analysis of its effect on particular fields of activities. Even in countries such as France, The Netherlands or Ireland which chose to conduct intermediate evaluation anyhow, the quality has improved, the evaluations being more focussed and usable for programmes improvement. This approach will not change from 2014 onward. What changes is the compulsory requirement to produce an evaluation plan early in the programming period and to conduct, at least once in the programming period, evaluations assessing "*how the support from the CSF Funds has contributed to the objectives for each priority*" (article 49).

These compulsory impact evaluations should favour different approaches, use of non-experimental methods where priorities are built on simple interventions that lend themselves these techniques, use of theory-based approaches where the programmes support complex interventions. The broad requirements of Article 49 (contribution to the policy objectives of each priority), should actually favour theory-based designs as most priorities build on different interlinked interventions. However we should be aware of the risk that the perceived purpose of these evaluations is only accountability omitting their learning dimension for policy making, with evaluations limited to quantitative estimation of the effects of different interventions, at best using qualitative techniques as an "add-on" to illustrate quantitative findings[15] without rooting the different methods in a rigorous evaluation design.

Another potential room for theory-based evaluation is policy experimentation. Experimentations were institutionalised in 2000-2006 through Regional Programmes of Innovative Actions dedicated to pilot innovative activities. This approach was mainstreamed in the current period however very few programmes if none took it up, Managing Authorities being reluctant to fund experimentations perceived as financially risky and administratively complex. In the coming period, Urban Innovative Actions (article 9 ERDF Regulation) will promote experimental approaches and solutions in the field of sustainable urban development at the initiative and under management of the European Commission. Managing Authorities can support local development strategies combining different kind of interventions which also would be good cases for applying theory-based approaches.

What are we doing to support theory-based approaches?

The weaknesses of the approaches and methods in ERDF evaluations are not a surprise: evaluation culture is still developing as are evaluation techniques and design, and evaluations need to adapt to an evolving regional policy. Articulating a good programme theory of change shared by stakeholders is challenging and requires competencies that programming teams often lack. A recent study commissioned by the Department of International Development analyses similar weaknesses in UK

[15] As was often used Case studies in the evaluations analysed in the Technopolis Group study, 2012.

Development programmes and their impact evaluations, underlining in particular the rare use of approaches linking interventions with outcomes such as Theory Based, Case-Based or Participatory approaches (Stern et al., 2012).

We encourage Managing Authorities to test new methods in their evaluations and allow for a combination of methods in their Terms of Reference. This can bring several benefits to the work: strengthen the conclusions by "triangulating" the outcomes through different approaches (this may also result in two contradicting analyses, making the evaluators refine their evaluation design) ; bring another dimension to the analysis, from the "what works" to the "how does it work"; improve analytical tools (use quantitative techniques to prepare the ground for - or challenge - qualitative work, or use qualitative fieldwork to lead quantitative work in useful directions); and in general terms, favour the techniques that are best adapted to answer the different evaluation questions according to the programme attributes[16].

Our aim is not to multiply techniques which would end with costly evaluations and bring no real policy added value. This mix of methods needs to serve rigorous evaluation designs driven by programme theories at the outset. The evaluation should start with the underlying theories of a programme and select the most relevant methods to elicit and test them against the intended or observed outcomes and examine the role played by the intervention and by other influencing factors. This relevance also depends on the evaluation budget and time constraints.

Another message to both commissioners and evaluators is to reflect on the targeted audience and on the use/usability of the evaluation. According to culture or background, evaluation "receivers" may be more sensitive to quantitative or to qualitative assessments of outcomes. Targeting a wide audience will request to develop some narratives with both quantitative and qualitative evidence to convey the main evaluation findings more easily.

The evaluation community - an essential partner in this step change

Over the last thirty years, advocates of theory-based approaches occupied the theory space promoting different labels for Theory based evaluations: Theory driven evaluation (Chen,1997), Contribution Analysis (Mayne, 2001), Elicitation Method, Impact Pathway Analysis (Bouthwaite 2004) or General Elimination Methodology (Scriven, 2008) to name some of them. This counterbalanced the increasing space occupied by experimental evaluations and their various techniques in the wake of the push for "evidence-based policy". The time has come to occupy the practice field: creating examples of convincing and useful theory based evaluations, proposing solid evaluation designs mixing quantitative and qualitative techniques, defending them by demonstrating their usefulness for policy making is the only way to convince evaluation commissioners.

[16] We understand "attributes" as in Stern et al.: intentions, activities and assumptions as well a technical contraints such as data availability, timing etc.

We need a critical mass of good quality evaluations, having proved their utility within reasonable costs, to raise awareness and promote exchange of practices. This could also allow for meta-analyses or "systematic reviews" giving insight into the necessary conditions for mechanisms that make an intervention work to hold throughout a variety of contexts.

A starting point is for evaluators to contribute to strengthen EU regional policy programmes' theories. The coming *ex-ante* evaluations are opportunities that should not be missed. In our guidance on *ex ante* evaluation, we assigned evaluators a central role to raise the quality of programmes design, looking both at assumptions and evidences supporting the theory, indicators and data sources, financial and human resources, including advising on future evaluations and their methods. The risk we anticipate is that, as often is the case in current programmes, theories will not be sufficiently explicit, sometimes confined into a logical framework which does not spell out the conditions under which the programme success is assumed. In particular, Managing Authorities may not sufficiently analyse the context of the intervention including other factors contributing to the outcomes (other interventions or policies, socio-economic trends etc.).

Evaluators should defend their position when answering a call for tender or working with evaluation commissioners. Sufficient time and effort should be allocated to analyse the programme theory and to refine the evaluation questions in interaction with the commissioning Managing Authority. These questions are often too broad or not realistic given the timing of the evaluation, sometimes they are not relevant. Evaluators need to propose alternatives based on the programme theories and the context in which they operate. Proper literature reviews and some preliminary fieldwork can provide insights to important factors that should be explored thus helping to generate useful hypotheses (H. White, 2008, 2009).

Managing Authorities rarely have the capacities to define the best approach and techniques, especially where the object of the evaluation is a complex development programme. To be safe, they tend to privilege the methods they understand, i.e. literature review, interviews, focus groups, case studies, and to avoid less "accessible" methods such as network analysis or counterfactual approaches. Some plan a step in the tendering process for applicants to present and discuss their proposed approach before the selection decision[17]. This opens up an opportunity for evaluators to convince on the relevance of their proposed approach, based on their experience and on an analysis of the programme's attributes. Key design decisions for complex programmes include "*how to decide on the unit of analysis when there are many related interventions taking place simultaneously and likely to influence each other; and when to conduct an evaluation of a programme that is long term, 'emergent' and unpredictable* " (Stern et al., 2012 p.64).

EU regional policy evaluations are rarely pure implementation or pure impact evaluations. Most are commissioned both for accountability and learning purposes.

[17] See for example a 2011-2012 Belgium (Wallonia) evaluation of ERDF co-funded actions to develop and exploit innovation potential in the ERDF programmes 2007-2013.

This added with the recent push toward mixed methods should encourage their uptake. However the responsibility of designing a coherent and rigorous evaluation design lies on the evaluators' shoulders. Mixing methods should not impact on evaluation costs, as budget and time constraints are not likely to disappear over time. Rather, the evaluators should propose (and defend) evaluation designs using the best of each family of methods to meet the evaluation expectations at reasonable costs and given the programme attributes.

Our combined efforts should aim to create a body of evaluations pragmatically combining methods in an as rigorous way as needed to answer the client evaluation questions. On this basis we could carry out systematic reviews supporting policy making based on evidence. The weaknesses of the current EU regional policy programmes (no explicit programme theories, loose theory of implementation) and their evaluations (insufficient focus on theories, few impact evaluations, no information on contexts, lack of detailed information on the evaluation process) need first to be addressed. Then new approaches have to be designed to ensure the credibility of evaluations' evidence.

Quality assurance of approaches and method is a prerequisite for external validity. Chen (2006) rightfully underlines the "*great need for systematically developing mixed method 'use' strategies as well as establishing its own standards and criteria for assessing the method use*". The DFID report discusses the possibility of common standards to be used across different designs and methods. The combination of Campbell standards and the realist evaluation approach proposed by Van der Knaap et al. (2008) is a path toward this direction that could be explored.

References

Astury Brad and L. Leeuw Frans: *Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation.* Evaluation, SAGE, 2010

Brulin Göran, Svensson Lennart: *Managing Sustainable Development Programmes, a learning approach to change*. Gower, 2012

Chen Huey T.: *A theory-driven evaluation perspective on mixed methods research*. Research in the Schools, Vol. 13(1), 2006

European Commission: *Guidance document: The programming period 2014-2020. Monitoring and evaluation of European Cohesion Policy – European Development Fund and Cohesion Fund. Concept and recommendations*. 2011
http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/wd_2014_en.pdf

European Commission: *Guidance document on ex ante evaluation, the programming period 2014-2020. Monitoring and evaluation of European Cohesion Policy – European*

Development Fund, European Social Fund and Cohesion Fund. 2012
http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/ex_ante_en.pdf

Hart Mark: *Evaluating EU regional policy, How might we understand the causal connections between interventions and outcomes more effectively?* Policy Studies, Vol.28, n°4, 2007

Leeuw Frans L.: *Theory Based Evaluation*, 2012
http://ec.europa.eu/regional_policy/information/evaluations/pdf/impact/theory_impact_guidance.pdf

Rogers Patricia J.: *Theory based evaluation: Reflexions Ten Years On*. New Directions for Evaluation, n°114, 2007

Rogers Patricia J.: *Using programme theory to evaluate complicated and complex aspects of interventions*. Evaluation, SAGE, 2008

Stame Nicoletta: *Theory-based Evaluation and Types of Complexity*. Evaluation, SAGE, Vol.10(1), 2004

Stern Elliot et al.: *Broadening the range of designs and methods for impact evaluations. Department for International Development*, Working Paper 38, 2012

Technopolis Group & MIOIR: *Evaluation of Innovation Activities. Guidance on methods and practices.* Study funded by the European Commission, 2012
http://ec.europa.eu/regional_policy/information/evaluations/guidance_en.cfm#1

Vanclay Frank: *Guidance for the design of qualitative case study evaluation- A short report to DG Regio*. 2012
http://ec.europa.eu/regional_policy/sources/docgener/evaluation/doc/performance/Vanclay.pdf

Van der Knaap F., Leeuw Franz L., Boagaerts S; and Nijssen T.J.: *Combining Campbell standards and the realist Evaluation approach: the best of two worlds.* American Journal of Evaluation, SAGE, 2008

Weiss Carol H.: *Theory-based evaluation: past, present, future*. New directions for Evaluation, n°76, 1997

White Howard: *Of probits and participation: the use of mixed methods in quantitative impact evaluation.* Nonie working paper N° 7, 2008

White Howard: *A contribution to current debates in impact evaluation*. Evaluation, SAGE, 2010

[i] Disclaimer: Marielle Riché works at Directorate General Regional and Urban Policy, European Commission. Views expressed in this paper are the sole responsibility of the author and do not necessarily correspond to those of the European Commission.