Evaluation Network Meeting

Lessons learned from reviewing Cohesion policy evaluations and from Helpdesk support to MAs

Terry Ward

Evaluation Helpdesk and Applica

11 December 2020

APPLICA

ISMERI EUROPA

Background

- As you all know, the Evaluation Helpdesk reviews all evaluations on Cohesion policy programmes produced in Member States and provides support on evaluation issues
- Helpdesk has so far reviewed over 1300 evaluations since 2015 on both present and previous programming periods
- And has provided support to many MAs in form of assessing evaluations undertaken or planned and advising on particular issues
- From these activities, number of lessons can be drawn on difficulties experienced in carrying out evaluations and on ways they can be improved
- General point – only small % of evaluations so far reviewed on 2014-2020 period have been on impact
- But % increasing and clear it will continue to do over next two years – these are main focus here

General lessons

- From support visits – clear that MAs, or evaluation units in them, are eager to carry out good and effective evaluations
- Aim – not just to comply with regulations but to assess effectiveness of programmes, results produced and how can be improved
- Clear too that in many cases, difficulties of carrying out good evaluations lie in part with state of local evaluation market – with competence of evaluators available to undertake them
- Some onus on MAs to improve situation – to develop and expand market, by e.g.:
  - encouraging academic involvement
  - supporting evaluation programmes or courses in universities
  - informing evaluators of planned evaluations and what is expected from them
  - encouraging use of international evaluation experts
- Procurement regulations also a constraint but a gradually weakening one.

Importance of Terms of Reference

- ToRs at root of good – and bad - evaluations
- Essential to get ToRs right to get what is wanted from an evaluation – features of good ToR:
  - clear indication of objectives - what we want to know and why – and of priorities, so that evaluation suitably focussed
  - clear evaluation questions and limited in number – focused on main objectives, so that evaluation effort not overly dispersed
  - questions capable of being answered given data available or possible to be collected
  - clear indication of data availability and gaps that need to be filled
  - possible indication of methods expected to be used, but onus put on tenderers to specify methods intend to use to answer questions
  - clear specification of evaluator requirements, including expertise in policy area concerned (e.g. of energy, environmental or transport issues)
  - clear specification of selection criteria and weight attached to each

Choice of methods

- Evaluation design needs to be capable of answering evaluation questions
- For impact evaluations, approach adopted is critical - needs to be capable of distinguishing effect of measures examined from other factors potentially affecting outcomes – including developments in context
- Majority of impact evaluations reviewed consist of analysis of monitoring data plus survey and interviews of recipients of support and/or implementers
- Former intended to identify what happened, latter effect of measure concerned on behaviour or performance
- Representativeness of survey respondents not considered in many cases
- Often potential bias of information obtained not explicitly recognised and adjusted for – not corroborated by information from other sources and compared with behaviour or performance of non-recipients
- Account often not taken of possible recall errors involved in indicating situation before receipt of support and describing action taken after receipt
- Surveys + interviews can produce useful evidence but ideally when carried out at various stages of intervention – before and after in particular

Methods used

- Minority of evaluations use counterfactual and/or theory-based methods – designed to take account of other factors potential affecting outcomes
- Even fewer use cost-benefit analysis - typically for large projects but not only
- Essence of counterfactual evaluations is to identify effects of interventions
- Essence of theory-based is to identify how effects produced – to identify causal chain leading from an intervention to outcomes and underlying mechanisms involved
- But for many evaluations, combination of two is needed to examine why or how interventions are effective as well as what effects they produced
- Even evaluations focused only on effects produced, however, should also examine how they do so – to map causal chain so that factors involved can be explicitly included in analysis
- Good counterfactual evaluations should therefore start by describing theory of change to be sure of taking account of all relevant factors that could affect outcomes

Issues in application of methods

- For CIEs, common failing is to define a control group that is not sufficiently similar to recipients of support to indicate the counterfactual
- Or, in case of Difference-in-Difference method, not to demonstrate that pre-intervention behaviour and/or performance of control group is close enough to that of recipients of support to be representative of it after
- For TBEs, major weakness is failure to spell out theory in sufficient detail that causal chain + mechanisms can be examined satisfactorily
- In many cases, difficult to detect theory of change at all – often diagram presented with arrows indicating links in chain without explaining mechanisms involved
- Detecting evidence of causal chain requires detailed investigation through field work – or case studies – examining way an intervention works in practice and following through the processes at different stages
- But this often not carried out with sufficient effort – reliance instead on surveys and interviews that are not focused on the various links in chain and mechanisms connecting them

Efficiency aspect

- Commission's 'Better Regulations' include 'efficiency' as one of criteria for evaluating an intervention
- Up to now, none of evaluations reviewed have assessed efficiency satisfactorily
- Most do not attempt to do so – those that do tend to relate outcomes to expenditure without taking account of extent to which measure assessed is responsible
- But difficult to do so because of problems in distinguishing costs associated with a particular measure
- And more seriously, in being able to identify a benchmark against which costs per unit of outcome can be measured – i.e. being able to estimate cost per job created does not indicate whether measure is efficient or not
- Even comparing unit costs between measures often problematic because of difference in recipients, underlying situation and surrounding factors

Data issues

- Access to relevant and reliable data is critical an evaluations to produce robust results
- An efficient and complete monitoring system to track the output and results of programmes is a key aspect in this regard
- But data also required on developments which potentially affect outcomes of measures or needed to carry out counterfactual analysis – i.e. to define a control group
- Administrative data tend to be the most reliable source, but few evaluations reviewed make use of them
- This not necessarily because such data are not available but because they are not accessible – primarily due to confidentiality issues
- These can often be resolved by anonymising data before being provided – but typically requires negotiating with authority responsible and forward planning
- 'Big' data also becoming increasingly available – used (on traffic flows) by two evaluations of transport investment (in EE and PL) but not in others

Timing

- Many impact evaluations reviewed carried out too early to reliably assess effects of measures because insufficient time elapsed for them to become evident
- General issue – time when an impact evaluation undertaken rarely ideal – too early for full effects of measures to materialise or too late to affect policy
- From evaluation perspective – longer time has elapsed, harder to isolate effects from other factors, but still need sufficient results to be evident
- Many evaluations reviewed assess estimated effects rather than actual since latter yet to materialise – limited usefulness, except in *ex ante* sense
- To overcome this can make use of evaluation findings on earlier period if measure not changed much – rarely done in practice and could be exploited more
- Can also investigate initial links in causal chain defined by theory of change to see whether predicted outcomes evident in practice – gives some confidence ultimate effects will also materialise, or not
- Two can be combined

Importance of forward planning

- A major lesson to be drawn from Helpdesk activities is importance of forward planning - to identify evaluations to be carried out, their timing, main questions to be asked, methods to be applied and data to be used.
- All MAs required to produce evaluation plans at start of period – encouraged forward thinking but most did not go far enough
- Most stopped short of identifying data required in enough detail to ensure they were collected as programme undertaken and available when needed
- Ideally, data needs and evaluation activities more generally, should be planned when policy measures initiated, as an integral part of programme design
- But for this to happen requires evaluation to be taken seriously by policy-makers
- Requires that evaluations produce findings considered useful for improving policies - which requires in turn that they use the most suitable methods and most relevant data …
- which means forward planning.

Thank you

for your attention

APPLICA

ISMERI EUROPA