STI policy evaluation
focus on the impact evaluation

Laurent Bach
laurent.bach@unistra.fr
BETA, University Strasbourg/CNRS

• A few definitions
• The key question of additionality
• Effects, outputs and impacts
• Some examples
• New challenges for an "ideal" evaluation scheme?
1. A few definitions

EVALUATION:

“An evaluation is an assessment, as systematic and objective as possible, of an ongoing or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision making process of both recipients and donors.”

(OECD 1987)

Analysis and assessment of goals, instruments and impacts (Meyer -Krahmer / Georghiou 1992)

EVALUATION OF S&T POLICIES:

Individual actors (researchers...)

Teams, labs

Organisations (PROs, HEIs, Tech Center, Science-industry intermediaries...)

Mechanisms / institutions (collaborative scheme, funding mechanism, IPR regulation...)

Companies

Not-For-Profit orga.

Projects / programmes

Policy

( Evaluation of local / wider S&T systems - Nat. or Reg. SI )
THE DIFFERENT LEVELS OF EVALUATION

E1 - E4: Relevance (content) and "quality" of conception (decision process)
E5 - E7: Implementation and results
E8 - E9: Efficiency and legitimization

Relevance (content) and "quality" of conception (decision process)

E1: Relevance of objectives
E2: Coherence of objectives
E3: Relevance and coherence of the "institutional arrangements"
E4: Coherence between objectives and institutional arrangements
THE DIFFERENT LEVELS OF EVALUATION

Implementation and results

E5: Programme management (cost-timing-quality...) = monitoring

E6: Effects / outputs / impacts = effectiveness

E7: Match between effects and objectives = efficacy (1)

Efficiency and legitimization

E8: Match between objectives - institutional arrangements and effects-outputs-impacts: do the same / do better another way = assessing adequate funding, management, contractual behaviour in order for objectives to be achieved in a cost-effective manner = efficiency
THE DIFFERENT LEVELS OF EVALUATION

Efficiency and legitimization

**E9**: Ex-post relevance of objectives, given the results of the other evaluations = assessing whether initial objectives are still valid in the light of evolving RTD, societal and environmental conditions = efficacy (2)
2. The key question of additionality

ADDITIONALITY and related concepts

What difference does State intervention (i.e. policy) make? 
<= What will happen (has happened) over and above what will happen (would have happened) anyway?

Does this difference justify State intervention?

= comparing situation «with policy» vs situation «without policy»

- efficacy (1) : with policy vs objectives
- efficiency : with policy vs with an alternative policy aiming at similar objectives
- opportunity cost : with policy vs other use of public funds

Policy Rationales, Evaluation Rationales and Additionality

- Neo-classical / standard approach (input-output approach)

  Market failures

  Input additionality
  whether the public action adds to, or substitutes for the agents inputs (usually financial) crowding out effects, displacement etc

  Output additionality
  whether the same/different results would have been obtained without policy action

Counterfactual -> substitution/crowding out - additive effects

State : «out» of system Normative approach
(Evaluator is independant)
Policy Rationales, Evaluation Rationales and Additionality

- Innovation system approach: System failures
- Evolutionary approach: Selection failures
- Knowledge-based approach: Knowledge/learning failures

**Behavioural additionality** (incl. Network additionality ?)
- the differences to the agent’s behaviour following the policy action
- the persistence of differences to the agent’s behaviour beyond actions carried out under policy
- the differences in network structures and actors’ position in structures

**Cognitive capacity additionality**
- whether the policy action changes the different dimensions of the cognitive capacity of the agent.

Counterfactual -> complementary, multiplicative, catalytic effects etc

State: part of the system

(Normative approaches? 
Evaluator is in the system)

Additionality in FP5 assessment (Polt-Streicher, 2005)

What would you have done if the project had not received FP5 fundings?
(n=939 participants)

- 57% would not have undertaken the project in the absence of EU funding
- 36% of the remainder would have gone ahead:
  - On a smaller scale: 29%
  - Less partners: 23%
  - Less ambitious objectives: 15%
  - Lower expectations of profit: 14%
  - More national partners: 12%
  - A longer time-scale: 11%
  - Less funds: 7%
  - 10% would have replace EU funds with their own funds
  - 20% would have replace funds with other external funds

- 13% would have done at least one thing “better” without EU funds
A BASIC VIEW ON (output) ADDITIONALITY

Approaches to Additionality Measurement

How to approach the "Without pp case"?

- Null hypothesis:
  - before ("without") / after ("with") comparison
  - baseline: "without" = same trend as before

- Counterfactual:
  - treatment group ("with") / control ("without") group
    - randomized field experiments (gold standard OECD)
    - quasi-experiments
  - “fictitious case” ("invent" alternative history)
How to approach the "Without pp case"?

• Existence of potential control group
• Scale (possibility of seizing control group)
• Possibility of controlling variables
  • complexity of events/causalities
  • spillovers between control and treatment groups
  • moving targets
  • ...
• Time scale

3. Effects, outputs and impacts

Economic evaluation vs economic effects, Direct vs indirect effects, short term vs long term effects, occasional vs steady effects, etc

• Competitiveness and exploitation/market-related effects (products, prototypes, patents, processes, interfirm technology transfer...)
• Individual and organisational learning effects (such as finding of new partners, networking, building of research infrastructure) and behavioural changes (such as increased collaboration skills)
• Influencing norms and standards
• Structural effects (firm size, regional distribution, trusts, cartels...)
• Externalities : generic knowledge, RD flows, system and network externalities, impact on the environment, health, education ...
• Cohesion, diversity (in case of international programmes, especially European programmes)
• Contribution to skills and research manpower

(Meyer-Krahmer / Georghiou 1992)
One reference model: outputs - (outcome) - impact

Table 4: RTD Outputs and Impacts

<table>
<thead>
<tr>
<th>Scientific &amp; Technological Output</th>
<th>Scientific &amp; Technological Impacts</th>
<th>Economic Impacts</th>
<th>Social Impacts</th>
<th>Policy Impacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papers</td>
<td>New Knowledge</td>
<td>Economic performance</td>
<td>Quality of life</td>
<td>Economic &amp; industrial development</td>
</tr>
<tr>
<td>Patents</td>
<td>Exchange of Knowledge</td>
<td>Industrial competitiveness</td>
<td>Social development &amp; Services</td>
<td>Follow-on Projects</td>
</tr>
<tr>
<td>Prototypes</td>
<td>Culture of Collaboration</td>
<td>Organisational Innovation</td>
<td>Control &amp; care of the environment</td>
<td>Regulatory Change</td>
</tr>
<tr>
<td>Products</td>
<td>Network Formation</td>
<td>Employment</td>
<td></td>
<td>Contribution to policies</td>
</tr>
<tr>
<td>Processes</td>
<td>Scientific Reputation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Services</td>
<td>Community Development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge &amp; Skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Codified - tacit ?
Shared - distributed among actors ?
Public - "club" - private ...

From ToolBox Report 2002

Outputs vs
- changes in routines
- behavioural changes
- structural & organisational changes

internally to actors - between actors
Outputs, changes in system.......... (outcomes).......... -> Impacts

**indications**

- PROPAGATION of effect (short term -> long term, participants -> non participants, micro -> meso -> macro)
- SEPARABILITY among factors generating effects (project fallacy - timing and scope, cumulativeness -, complementary assets for innovation, between partners if collaborative projects)

Need models, theoretical background: domain impacted (science, technology, economy, social, environment, policy...), actors, relation between actors and propagation modes, time horizon, positive vs negative effects <=> reference

Ex. market, evolutionary schemes, network approach...

**Dominant models (1):**

NO MODEL !! List of indicators "à la Prévert"
Dominant models (2): linear model / input-output based

- FONDAMENTAL RESEARCH
- APPLIED RESEARCH
- DEVELOPMENT
- MARKET INTRODUC. Production Sales

Scientific publications
Scientific and Technical publications, patents

methods, models, data base, reports, lab notebooks, ...

Technical Publications Patents, prototypes, pilots

Products, processes, services
+ Sales, income...

Effects on clients, competitors, suppliers, partners, society as whole...

Dominant models (3): « upstream » (scientific production, technology...)
Development of STI indicators

- Publication - citations (“bibliometrics”)
- Patents/licencing
- Grey literature e “grise” (reports, notes...)
- Objects/artefacts
  - pilot, prototype
  - products, process
  (=> “technometry”)
- Services
- Methods, models, standards, data base
**Dominant models (4):**

« upstream » (scientific production, technology...)

**Peer review**

Evaluation process based on judgements formulated by independent experts = independent of the work being reviewed.

US General Accounting Office (GAO) : “a process that includes an independent assessment of the technical, scientific merit of research by peers who are scientists with knowledge and expertise equal to that of the researchers whose work they review.”

**Goals:**

- review of the research results submitted for publication in scientific journals *Journal manuscript peer review*
- allocation of research funding (generally ex ante evaluation) *Proposal peer review*
- assessment of the quality of research conducted by Research operators, e.g. Universities (generally ex post evaluation) *Evaluation peer review*

**General organization (Birta - Lung 2006):**

- Paper-based reviews/or mail-in reviews : written opinion on the work; no relation between reviewers (Journal pr)
- Panel of experts : single evaluation, generally focuses experts (evaluation pr)
- Standing committee : set term, number of reviews of different projects/programs, generally broad range of expertise covering many subject areas (proposal or evaluation pr)

**General problems-limits** (see also limits of indicators and method used) (Birta-Leung 2006, Reale 2006):

- subjectivity of the judgements
- absence of specific expertise of peers
- bias in favour of prestigious institutions, prominent researchers and research groups based on reputation or other recognitions given to the work
- interdisciplinary bias
- leakage of ideas (deliberate or subconscious incorporation of ideas proposed in work being reviewed into reviewer’s own work)
- conservatism (reduced likelihood of presenting radical ideas if it is believed that they will have less chance of being funded)
- neutrality/impartiality (difficulty in assembling a review panel that is both knowledgeable and neutral i.e. no judgment according to personal or institution interest)
- transparency and rationality (rationale of process, internal rules, objectives, indicators, criteria and evaluation methods etc not fully known, acknowledged, respected by all stakeholders - reviewers, reviewees, policy-makers etc)
But DIFFICULTIES related to the very nature of R&D activities...

uncertainty, long term, spreading, multiplicity of actors and factors...

...to the scope of effects ...
which actors, intangible effects ...

...to the type of STI policy
fundings, IPR regulation, networking

<= linear model- "input/output" - private vs social rate of return

Dominant models (5) :

« Downstream » (market...) 
Cost-benefit analysis
Ex. :
- products / process services derived
- sales of R&D services, tests, technical assistance
- revenues from the commercialisation of R&D results (technologies, patents,...)

(use public (and private !) for smtg else ...
Laissez-faire

Public lab.
Public support
Public support?

Private threshold

Social rate of return

High

Low

< 0 (pollution...)

Low

Public support

Private rate of return

High

Use public (and private !) for smtg else ...

Stop the privaters ?!
Neo-classical microeconomic framework for the evaluation of R&D project:

Private / public rate of return and spillovers (externalities)

Adapted from Jaffe

Dominant models (6):
ECONOMETRIC MODELS (I/O based)

=> firm, sector, macro levels
output = f (various inputs possibly incl. those with public origin, structural variables, dummy for participation to programme)
**Dominant models (7):**

**INTER-INDUSTRIAL FLOW MODELS (I/O based)**

Official statistics on flows of goods and services between sectors: 1 € in targeted sector => x € in sector i, y € in sector j etc

**Direct effects:**
Economic activity for contractors and direct suppliers of contractors

**Indirect effects:**
Economic activity for the whole economy via supply-chain links (suppliers of suppliers of...)

**Induced effects:**
Economic activity via employees’ expenses

---

**... AND SOME OTHERS**

- **Network based approaches:**
  Social networks: increase of number of links, centrality, connectivity... => **growing emphasis**
  Techno-economic networks (CSI group)

- **Multidimensional approaches**
  ex ESAC approach (Brazil)

- **Logic model approach**
  linking goals-instruments-impact-evaluation targets

- **Knowledge-based approaches**
  BETA approach (to some extent)
Example: networks in EU collaborative programs

<table>
<thead>
<tr>
<th>Source</th>
<th>Focus</th>
<th>Network considered</th>
<th>Network dynamics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malerba et al. (2006a)</td>
<td>Mapping of networks and identification of hubs in ICT and IST</td>
<td>Network of FP participants and networks generated by other collaborative initiatives</td>
<td>No ex ante, ex post comparison</td>
</tr>
<tr>
<td>Wagner et al. (2005)</td>
<td>Investigation of overlaps and interconnections among networks</td>
<td>Analysis of the relative position of actors in the networks</td>
<td>Not the same nodes</td>
</tr>
<tr>
<td>Malerba et al. (2006b)</td>
<td>Geographical dimension of FP networks and ERAnets</td>
<td>Network of FP/ERAnet participants</td>
<td>No</td>
</tr>
<tr>
<td>Malerba et al. (2010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perez (2010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breschi (2008)</td>
<td>Collaborative activity of FP participants as compared to the general population</td>
<td>Co authorship</td>
<td>No</td>
</tr>
</tbody>
</table>

Note: snapshots of networks generated around/by EU prog but no study of the impact of EU prog on networking activity

Adapted from C. Hussler - Beta 2011

Figure 21. General Framework for Goals, Instruments, Activities, Impacts and Indicators
BETA approach

RD programmes (RD content - project characteristics: limited in goals, time) especially collaborative ones

Ex. FP programmes - RD collaborative projects (public co-funding)
    European Space Agence programme (public procurement)
    Material Ireland (private/public funded technical centers)
    French national programme for SMEs (refundable grants)

Sampling (participants / projects)

Micro evaluation and sum up of results at sample level

Data collection based on direct interviews

• Effects: for participants only
  Direct effects: fit with objectives
  Indirect effects: beyond objectives - various learning effects affecting other activities than those related to evaluated activities
  - scientific and technological
  - networking, relational
  - organisational changes (internal)
  - methods, modes of management
  - reputation
  - critical mass

• Identification of effects / learning processes (intra & inter partners)

• Quantification:
  - products/services sales, cost reductions, new research projects, revenues from licenses etc
  - "value" of critical mass
  - "coefficient of influence": very seldom 100%, minimal estimates
4. Some examples

- EUREKA: [http://www.eureka.be](http://www.eureka.be) CSE (Continuous and Systematic Evaluation) + new evaluation study 2006
- USA: ATP programme [http://www.atp.nist.gov](http://www.atp.nist.gov) implementation of 1993 GPRA (Government Performance and Results Act) and 2001 PART (Program Assessment Rating Tool) => effectiveness, efficacy (type 1= goal attainment)
- Various national experiences in :
  - Conf. OECD Paris 1997 Conference
  - Tokyo 2002 Conference
  - Kuhlmann/Shapiro 2003 book
  - ASIF 2003
  - Toolkit 2002
  - Microeconometric approaches (see Klette and al.)
  - Evaluation of basic research (2 conf in Vienna: [http://www.fteval.at/conference06/](http://www.fteval.at/conference06/))
- EUropean FOrum on Research and Development Impact Assessment

+ see references (reading list)
5. New challenges for an "ideal" evaluation scheme?

**Trends in S&T policy making**

- Multi-levels multi-agents decision
- Variety of goals
- Variety of Inst. Arrang.
- Coupling with other policies
- International collaboration / integration (EU)
- Flexibility / adaptive - learning policy makers
- Privatization / contractualization
- Development of competition-based programmes

**Trends / challenge for evaluators**

- Complexity/Separability
- Multiple stakeholders
- Coherence of objectives
- Weight of different dimension of the evaluation
- Handling/using the results
- Evaluation of policy mix
- Mix of « evaluation cultures »
- Evolving schemes
- Evaluation of shifts
- Legitimacy, project fallacy

---

**Trends in research activity**

- Globalization
- Solving problem orientation
- Interdisciplinarity
- Cooperation
- S-I linkages
- Increasing knowledge content
- IPR regulations
- Devpt of ICT

**Trends / challenge for evaluators**

- International dimension
- Benchmarking
- Bias toward short term market-type output
- Limits of peer reviews
- Separability/Complexity
- Network evaluation
- Evaluation of knowledge/competences/capabilities/capacity
- Limits of bibliometrics
Towards a « good » evaluation system ?

Complexity : no single method ; multi-interdisciplinarity /quantitative and qualitative; WITH alignment between (implicit or explicit) theoretical basis about which effects are to be evaluated

Guaranteeing the scientific value of evaluation methods (Robustness, Repetability, Appropriability, Transparency, Independance of the evaluators, Confidentiality, Sampling...)

Balance between systematic / standardized /simple approaches and exploratory studies

=> systematic: Evaluation vs monitoring, time fit between R&D activity and evaluation, Availability of data, Evaluation “fatigue”, Cost of evaluation

Ensuring that evaluation takes place on a programmed and properly resourced basis

Providing « easy to understand » results while avoiding meaningless list of indicators / scoreboards

Providing a mechanism for feedback of the results into policy making (learning policy maker)

Interactions between academics, practitioneers, policy makers and research actors for better understanding of scope, relevance and needs => evaluation as a « social process »
Evaluation as a « social process »
(from L. Georghiou)

- **Motives / interest relates to the actors:**
  - Those « being » evaluated
    - justification/legitimation
    - learning at operational level
    - gain new supports for public sources
  - Those who are the audience of the evaluation
    - accountability
    - resources allocation
    - learning at policy level (pro-active evaluation)
  - Those performing the evaluation
    - academic interest
    - consultant business

  ⇒ Make purpose and context-dependancy clear before choice of approach

---

Co-evolution of policy rationales / tools
and evaluation rationale/tools

- 1970s modification of peer review to extend criteria
- 1980s interest began in collaborative R&D programmes
- 1990s rise of:
  - performance indicators
  - emphasis on knowledge transfer indicators
  - commercialisation of research
- 2000s interest in:
  - evaluation of system capabilities eg national or regional systems
  - aggregate or interactive effects of policies (“policy mix”)
  - effect of “soft” policy tools such as foresight
  - Strategic and persistent effects on firms of public support
    (“behavioural additionality”)
  - Development of "logic models" linking goals-policy tools-evaluation targets
  - "New instruments" : NoE, IP, ERA-Nets, JPI, RI ... => new challenges
Evaluation of the EC STI policy

- 1995-2006: quite simple
  - Yearly monitoring
  - 5-year assessment looking back (past and ongoing FP) and forward (future FP) (high-level panel, interviews, studies etc) 1992-97, 95-99, 99-2003
  - Specific impact assessments: Energy, NMP...

- FP 7: less clear
  - Ex-post evaluation of FP 6 in 2009
  - Overall ex-ante evaluation of FP 7
  - Annual monitoring (indicator-based assessment of programme implementation)
  - Progress report in 2009
  - Interim report in 2010

- A lot of evaluation studies launched here and there (ex largest on FP: by DG Enterprise!)

=> Criticisms from the European Court of Auditors 2007-08