

**Doc. 8**

**CSS Commission on “The evaluation of the effects of public policies”**

**INTERNATIONAL WORKSHOP ON IMPACT EVALUATION  
PRACTICE AND PROSPECTS**

**April 13-14, 2007**

**Verduno, Italy**

**Background, issues for discussion and outline of the programme**

*(AM and UT, 23.03.07)*

**1. BACKGROUND**

The Foundation “Consiglio italiano per le Scienze Sociali” (*Italian Council for the Social Sciences*) recently set up a Commission on “The evaluation of the effects of public policies: methods, practices and perspectives”. The Commission comprises 11 members, mixing university-based researchers and high-ranking civil servants, and is co-chaired by Ugo Trivellato and Alberto Martini. (Members of the Commission are listed in *Annex 1*.)

The main motivation for setting up the Commission is the lack of a rigorous impact evaluation practice in Italy, which has several plausible explanations: policy debates characterised by extreme political partisanship; resistance to the introduction of a result-oriented culture in the Italian public sector; last but not least, the absence of quantitative training among most civil servants.

The purpose of the Commission is to produce a “White Paper”, with the aim of (i) clarifying the role of impact evaluation in public policy-making; (ii) identifying the main obstacles to the use of rigorous evaluation practices in the Italian context; and (iii) making recommendations for a gradual adoption of such practices (with a focus on welfare, education, labour and industrial policies).

**2. THE WORKSHOP WITH INTERNATIONAL EVALUATION EXPERTS**

An important step in the activities of the Commission will be a closed door workshop with international evaluation experts, to be held on April 13-14, 2007. The experts who have agreed to participate are listed in *Annex 2*.

The goal of the workshop is to gain insights into the processes by which other countries have implemented the evaluation of public policies and programmes. Our emphasis is on *impact evaluation* (i.e., estimation of causal effects of treatments on outcomes), while we recognise the importance of – and the necessary interaction with – other phases of a broader evaluation process, particularly with *implementation analysis*.

We look for contributions on the various aspects of impact evaluation: from the very definition of impact of a policy, to the critical design issues that must be tackled in order to produce credible results; to what happens when the “correct” impact evaluation approach

meets the real-world policy-makers. We are looking especially for insights into the institutional aspects of impact evaluation: when and why policymakers demand it; how the design of the evaluation is affected by data and resource availability; how evaluators are chosen; when and how impact evaluation results are utilized.

We do not expect participants to do a formal presentation at the workshop, rather to interact in a discussion that will evolve around a limited number of issues. In the following section we present a set of questions for the workshop participants to think about. We encourage participants to provide us, if possible, in advance of the workshop, with **short written answers** to the questions they think are more relevant.

**More importantly, we kindly ask workshop participants to give us in advance references to evaluation reports or journal articles directly relevant to the issues under discussion, with special attention to what they think are the best examples of impact evaluations conducted in their respective countries.** Particularly if they intend to refer to these examples during the discussion, it would be important for us to have the references in advance of the workshop.

**Please direct the references, including direct web links, to Luca Mo Costabella (e-mail: [lmocostabella@prova.org](mailto:lmocostabella@prova.org)), preferably by *February 28, 2007*.**

### **3. ISSUES FOR DISCUSSION IN THE WORKSHOP**

The following is a list of questions intended to offer guidance and structure to the discussion during the workshop. The questions focus on the motivation, the design, the data, the utilization of evaluation and the choice of the evaluator.

#### ***A. Who does request impact evaluations and why?***

**We are interested in learning about the process by which policy-makers request impact evaluations, and about the incentives and motivations they have in making (or not) such requests.**

**The following questions refer to what motivates the demand for impact evaluation.**

1. How often does the motivation for requesting an impact evaluation stem from a genuine desire to *improve a policy already in operation*, by identifying “what works” (and what does not)?
2. How often is evaluation used to *test in advance the effectiveness* of a policy by implementing it as a pilot test? How common is the use of what in the US are called *demonstrations*?
3. How often does the motivation for evaluation stem from a more generic call for *accountability for results*? In this case, is the notion of “impact” used by policy-makers still equivalent to that of “causal effect”, or does it entail a broader (multifaceted, possibly vague) meaning?

4. How often the motivation is of a *symbolic* nature, *i.e.* the call for evaluation is a way for policy-makers to reaffirm their commitment to the underlying policy issue, although they have no strong interest in the actual evaluation results?
5. Whatever the motivation, how common is to include an impact evaluation in the *design* of a programme, or at least to plan the evaluation *before* the programme is implemented?

**The following are questions on the type of institutional actors involved in demanding impact evaluations.**

6. How crucial is the presence of “*checks and balances*” between different branches and/or levels of government in determining the demand for evaluation? How often are impact evaluations demanded directly by the legislative branch while exercising its oversight on the executive branch?
7. How often does the demand for evaluation stem from the interplay between different *levels of government*, where one level funds a policy or programme that is going to be implemented by another (usually lower) level?
8. Within the executive branch of government, the policy-makers requesting evaluations tend to be *politicians* (or *political appointees*) or rather high-ranking *career officials*? How do motivations and incentives differ between the two groups?
9. What is, on average, the awareness of the practical and conceptual difficulties in making causal inference among those policy-makers who specifically request *impact evaluations* ?

#### ***B. What does influence the design of impact evaluations?***

**We are interested in understanding the process by which the evaluation design is chosen and which constraints weight more heavily on the choice. We are less interested in the technical aspects of evaluation design.**

**The following are broad questions on the design of impact evaluations.**

1. For designing an impact evaluation, how critical is that the *objectives* of the policy or programme are clearly stated in advance? How critical is that they are expressed as *numerical targets* (*e.g.*, x% reduction in the rate of...)?
2. In defining the “treatment” whose effects are to be evaluated, how much attention is devoted to identify the treatment *actually received* by the beneficiaries, as opposed to the statutory treatment? Is *implementation analysis* used in identifying the treatment actually received?
3. In general, how often is an implementation analysis conducted *in conjunction* with an impact evaluation? In which sense can the two be “integrated”, given the different methods by which they have to be conducted and the different units of analysis? Could you provide examples of such an integrated approach?

4. How much attention is devoted to *understand the causal mechanisms* that generate the effects of the policy *vs.* to *estimate of the direction and size* of the causal effects? Which methods can be used to understand such mechanisms?
5. How much attention is devoted to the *participatory* aspects of evaluation? Is there an active involvement, in defining evaluation criteria and design, on the part of beneficiaries of the policy, directly or through their advocacy organizations? Should these participatory concerns be seen as crucial for *impact* evaluation?

**The following are more specific questions on the use of the experimental design.**

6. How much reliance is there on the *experimental design* for impact evaluation? Even when not utilized for ethical or practical reasons, is the experimental design still considered an ideal benchmark? How widespread is the use of non-experimental methods that try to mimic the logic of experiments (as in the matched samples design), as opposed to methods that are model-based (as in regression analysis)?
7. When the experimental design is used, is the random assignment mainly done with respect to *sites/areas* or with respect to *individual units* (persons, households, firms)? How are ethical objections overcome in the two situations?
8. When an experimental design is used, especially when random assignment is done with respect to individual units, what are the *operational procedures and practices* adopted to assure a sound implementation of the design? Are they reasonably successful in avoiding the potential threats to randomization coming from the behaviour of the service providers and of the treated and/or control units?

### **C. Where do the data for impact evaluation come from?**

**We are interested in the consequences of data availability for the design, the cost and more generally for the feasibility of impact evaluation. We would like to understand how each country positions itself along the *continuum* between a situation in which evaluation design is *data-driven* (*i.e.* the availability of data is taken as given and constrains the choice of methods) and the situation in which data collection is *design-driven* (*i.e.* the data is almost always collected *ad hoc* as an integral part of each evaluation).**

**Thus, the introductory question refers to the prevailing practices about the data used for impact evaluation. If possible, please provide us with examples of each type of practice.**

1. How often is the evaluation design to be *data-driven* and how often is the data collection to be *design-driven*?

**The following questions refer to the situations in which the data for evaluation are drawn from existing sources.**

2. When no primary data collection is conducted, are the data obtained mainly from *administrative sources* or from existing *household or firm-level surveys*? What are the constraints on the access to these data imposed by confidentiality concerns? How are these concerns addressed when the data are to be used for evaluation/research purposes?
3. How the data needed for impact evaluation fit into the *overall statistical information system*, particularly the component of the system based on *register data*? Are register data able to provide detailed background information on life histories of individuals and/or firms?

**The following questions refer to the situations in which the data are collected *ad hoc* for evaluation purposes.**

4. Are the data collected for evaluation purposes subsequently made available for *further use* to allow replication of the evaluation results? If so, is that done in terms of production of *public use files* or *via* some procedure of *restricted access* to qualified researchers?
5. Which share of the total evaluation costs is usually represented by the costs of data collection?

#### ***D. How are the evaluators chosen?***

**We would like to know the prevailing practices in selecting the evaluators and understand the trade-offs between different arrangements for conducting an impact evaluation.**

1. What is the prevailing practice in selecting evaluators? Does it favour “Evaluation Units” within the government agency in charge of the policy, or specialized government agencies, or outsourcing to private (or university-based) research organizations?
2. What are the dis/advantages of having the evaluation conducted by special “Evaluation Units” located *within* the government agency in charge of the policy? How long is the process for setting up such Evaluation Units?
3. What are the dis/advantages of having the evaluation conducted by *another government agency* (including public bodies in charge of the oversight of government activities: *e.g.* the Government Accountability Office in the US or the National Audit Offices in European countries)?
4. What are the dis/advantages of having the evaluation outsourced to *private research organizations* (or university-based research centres) through *ad hoc* contracts?
5. To what extent is such outsourcing motivated by the need for “independence” on the part of the evaluator, or instead by the need for specialized methodological expertise and operational flexibility in conducting the evaluation?
6. In the case of outsourcing, are competitive procurement procedures followed to select the evaluator? Are there difficulties in administering effectively a

competitive procurement for evaluation contracts on the part of government agencies?

**The following questions concern the relationship between the evaluator and the other actors.**

7. During the implementation of policy subject to an impact evaluation, is there a complete separation between the agency in charge of the policy and the evaluator (either a government agency or an external evaluator), or is there significant *interaction* between the two parties? If some interaction takes place, what are its typical features?
8. Does the interaction possibly involve also the beneficiaries of the policy (or their representative bodies – *e.g.* as in the case of industrial policies)? Do these interactions possibly affect the policy and the evaluation design itself, and in which direction?

#### ***E. How are impact evaluation results used?***

**We would like to have a general picture of the degree of utilization of impact evaluation results. We would like to understand if there are incentives or rules (or features of the evaluations themselves) known to increase the likelihood of utilization.**

1. Is there a positive relationship between the quality of the evaluation and the likelihood of it being used? In case there is such a relationship, which is the likely direction of causality? Better evaluation favours utilization, or is the prospect of utilization that stimulates better evaluations by motivating the evaluators?
2. Are there significant examples in which “interim” impact evaluations have been used to modify the implementation of a policy or programme? Are there significant examples in which “final” evaluations have been utilized in the decision to dis/continue a policy or programme or to alter its design significantly?
3. If there are no examples of such direct or instrumental use, is there evidence of a more indirect utilization? Of what Carol Weiss calls “enlightenment” function of evaluation<sup>1</sup>?
4. Are there some general rules or accepted practices for the dissemination of the results of impact evaluations? Do the media tend to utilize those results?
5. What is the *perception of the utility* of impact evaluation among policy-makers? And among the evaluators themselves? And among the media and the general public?

---

<sup>1</sup> “Governments don't often use research directly, but research helps people reconsider issues, it helps them think differently, it helps them reconceptualize what the problem is and how prevalent it is, it helps them discard some old assumptions, it punctures old myths.”

#### 4. OUTLINE OF THE PROGRAMME

The workshop will take place on *the 13<sup>th</sup> and 14<sup>th</sup> of April 2007* (a Friday and a Saturday), but the participants are expected to be in Verduno *for dinner on the 12<sup>th</sup>*.

*Verduno* is a small, nice town some 50 km south of Torino, in the middle of *Langhe*, the wine region *par excellence*. The workshop will be held at the *Castle of Verduno*. <http://www.castellodiverduno.com/>

The closed door *workshop will convey some 25 people* (international experts, members of the Commission and few other Italian experts). The workshop will deal with the issues listed in Section 3. We plan to have *four sessions on Friday and three sessions on Saturday*.

##### *Tentative schedule for the workshop*

###### **Friday 13**

- |             |  |
|-------------|--|
| 9.00 –10.30 | Brief introduction and opening remarks by the experts          |
| 11.00-12.30 | Topic A: Who does request impact evaluations and why?          |
| 14.00-16.00 | Topic B: What does influence the design of impact evaluations? |
| 16.30-18.30 | Topic C: Where do the data for impact evaluation come from?    |

###### **Saturday 14**

- |             |  |
|-------------|--|
| 9.00 –10.30 | Topic D: How are the evaluators chosen?          |
| 11.00-12.30 | Topic E: How are impact evaluation results used? |
| 14.00-16.00 | General discussion                               |

For the late afternoon of Saturday we will organise a visit to *Palazzo Madama*: it contains the whole history of the city of Turin within its walls; besides, presently it hosts the Exhibition “Alexander’s Heirs. From Seleucia to Gandhara”. Then we will move to the *National Cinema Museum*, located inside the *Mole Antonelliana*, and have the closing dinner there.

#### 5. PRACTICAL ARRANGEMENTS

The “Consiglio italiano per le Scienze Sociali” will cover travel and lodging expenses.

Workshop participants will stay at the Castle of Verduno from the 12<sup>th</sup> to the 14<sup>th</sup> of April. They will have an hotel booked in Torino for the night from 14<sup>th</sup> to 15<sup>th</sup> of April.

On the 12<sup>th</sup>, transportation to Verduno will be arranged from the main airports nearby (Torino Caselle and Milano Malpensa).

Invited experts are kindly requested to buy directly their plane tickets, at the most convenient fare. Experts flying from outside Europe might take a business class ticket if they need so, still at the most convenient fare. They will be reimbursed in a reasonably short time.

**Annex 1: Members of the CSS Commission on “The evaluation of the effects of public policies: methods, practices and perspectives”**

Ugo Trivellato (co-chair)	Professor of Economic Statistics, University of Padova.
Alberto Martini (co-chair)	Associate Professor of Economic Statistics, University of Piemonte Orientale.
Fabrizio Barca	Director, Ministry of the Economy; formerly Director of the Department for Development Policy, Ministry of the Treasury.
Manin Carabba	Centro Europa Ricerche; formerly Section’s President at the Italian National Audit Office.
Andrea Mancini	Director, Census Department, Italian National Statistical Office.
Enrico Rettore	Professor of Economic Statistics, University of Padova.
Antonio Schizzerotto	Professor of Sociology, University of Trento.
Paolo Sestito	Director, Research Department, Bank of Italy; formerly Economic Adviser to the Italian Ministry of Labour and Welfare.
Paolo Silvestri	Professor of Public Finance, University of Modena and Reggio Emilia.
Ignazio Visco	Vice-Director General, Bank of Italy.
Alberto Zuliani	Professor of Statistics, University of Roma “La Sapienza”; formerly President of the Italian National Statistical Office.

**Annex 2: Invited evaluation experts**

Marco Caliendo	Senior Research Associate, DIW, Berlin.
Peter Fredriksson	Director-General of the Institute for Labour Market Analysis (IFAU), Uppsala.
Denis Fougère	Research Director at CREST-INSEE, Paris.
Arianna Legovini	Senior Economist and Impact Evaluation Cluster Leader, Africa Region, The World Bank.
John Martin	Director for Employment, Labour and Social Affairs, OECD
Rebecca Maynard	Professor of Education and Social Policy, University of Pennsylvania; formerly Senior Vice President, Mathematica Policy Research, Inc..
Robert Walker	Professor of Social Policy, University of Oxford.