Evaluation and learning in the current crisis

Nicoletta Stame
University of Roma «Sapienza»
Australasian Evaluation Society Conference
Adelaide, 30 August 2012

Outline

- A general crisis, challenges for evaluation, remedies proposed

- Distributive policies and evaluation

- The EBP movement: its policy implications

- Learning and evaluation

A general crisis

- What kind of crisis: economic (financial, productivity), political (centralistic institutions), social (growing inequality, lack of confidence)

  >> Need for rigour, equity and development

- Failures ("nothing works"): what way out?

  – No more postponement of search for solutions

  – Discovering existing good practices, favouring innovation

learning

- evaluation familiar with single loop learning: correcting error according to theory in use
- need for double loop learning: correcting error by modifying organizational norms, developing a new conceptual framework.

>> learning to learn (Argyris and Schon)

Evaluation that favours development

>> developmental evaluation (Patton)

The role of evaluation in favouring development

- Surfacing society's latent energies: innovation, existing good practices
- Appreciate diversity: equity
- Build trust: improve democratic policy-making

What are suitable approaches?

Two policy scenarios

- In ordinary public policies, service delivery: how to resist linear cuts. With more evaluation? Better evaluation? Better use?

- In programs: the Evidence Based Policy movement: more robust evaluation for informed decision-making

What policy implications?

Ordinary public policies

- Distributive policies: "assistenzialismo", fragmentation, allocation for consensus, short-term perspective

- Ordinary spending: a normal budgetary process, allocation by functions

- Linear cuts (negative-distributive) imposed from the center down vs. a spending review based on evaluation (reducing waste, increasign productivity)

- Resistance to cuts based on the Minister's power, general directors' and down

No role for evaluation ? What role for evaluation?

Building trust

League tables, ranking, monitoring indicators. How to use them?

- - Helping the less performers to improve?

- - A premium for greatest improvement?

- - Transparent criteria: favour horizontal benchmarking

- - Link to causal anlysis of change (upward and/or downward)

Contributing to a more democratic policy-making

Setting standards:

- multiple sources of standards, pluralism in determining quality (Davidson)

- Horizontal benchmarking: different ways of tackling problems can be compared (Sabel)

Facilitating innovation

- recognizing innovations and helping strengthen them (facilitation instead of enforcement)

- Learning from success and «not punishing those who try, even if initial efforts are not perfect» (Perrin)

>> positive thinking approaches

- Appreciative Inquiry (Preskill and Torres)
- Success case method (Brinkerhoff)
- Most significant change (Dart and Davies)
- Positive deviance (Sternin)

Programs: The Evidence Based Policy movement

"nothing works": "what works?"

- provide sound evidence of what works by "robust" methods of impact evaluation (the evaluator as a methodologist)
- Policy-makers will take decisions based on evidence, not on ideology (learning function, instrumental use)
- What works can be scaled up

Only methodology failures?
What are its policy consequences?

Attribution.1

"Policy-makers want to know whether their intervention can be attributed a positive effect". Attribution = the intervention is the only and necessary cause of the effect.

Possible? Correct?

- Attribution more likely to show failures
- Most current policies are based on collaboration, partnership (Paris Declaration): avoid isolationism

Attribution.2

- Effects can be obtained in many different ways:
  - different «causal packages» (intervention plus contextual and policy factors)
  - Intervention as a «contributory cause» (with other interventions)
- Need to assess the contribution of partners: added value
  - allow for diversity and combination

"Robust methods" for evidence.

- A hierarchy of methods, with RCTs at the top and ethnographic studies at the bottom
- Interventions as treatment, the medical metaphor
- Social experimentation? Exploration vs. repetition

How robust are these methods?

RCTs are fit for simple causality, stable contexts, linear implementation, large n

That is to say: Not fit for most programs

With different questions (causality, explanation, equity) and complex attributes of programs (multi-site, multi-intervention etc.), other approaches better suited:

Approapriate approaches

Theory-based approaches, contributory analysis
- can answer why (theory testing, theory building)
- Eliminates rival explanations
- Able to deal with context

Comparative case studies, e.g. Qualitative Comparative Analysis
- Able to identify complex systems and compare necessary and sufficient factors

Participatory approaches
- Recognize people's agency as cause

What policy consequences?

- A suggestion to implement only those programs that can be evaluated by "robust" methods (Duflo)

- The risk of not being able to perceive what should be done to improve programs

Guidelines for a "correct" implementation?

Uniformity downward?

The principal/agent theory and distrust (moral hazard, adverse selection)

What is an implementation failure? Acting in an incorrect way, or acting in a different way?

- The risk of conformity and ritualistic compliance: control from above

- Not appreciating innovation and responsive adaptation: "discretionality" of street level bureaucrats and of entrepreneurial managers, responsible accountability.

Implementation and diversity

- Contexts, embedded situations

- Different trajectories (Woolcock): J curve (things go worse before getting better), step function (slow initial uptake, then rapid increase): pushing motivation at different points, and keeping it high

- Emergent trends (tipping points)

Should commissioners request specific designs?

Term of Reference that request specific designs and methods: RCTs, quasi-experiments, etc. An improvement?

- Designs should be appropriate to program characteristics and evaluation questions

- Designs should be agreed between commissioners and evaluators in a constructive dialogue, not imposed from above

Scaling up?

Is scaling up a reasonable goal? Uniformity upward?

- Generalization (to similar situations and circumstances) vs. transferability (to other situations, and adaptation by local agents).

  - Able to say "it worked there", but not "it will work for us, in other circumstances", which needs adaptation (Cartwright)

- Or: getting the same results with a different combination of mechanism and context? (Pawson)

Learning and use

The implication that having learned the "good lesson" (evidence of what works) then action will follow. True?

An instrumental use of evaluation?

– Or other types of use: cognitive, process?

– Ledermann: Context conditions (conflict, need for change) and evaluation results conditions (novelty, quality) for use

An engineering mode of learning ?

– Or an autonomy-respecting mode of learning?

An example: Learning in development

Ellerman: centralistic implication of direct aid interventions (the helper-doer relationship).

Cognitive and motivational dimensions:

- it is the centre that knows the solution and then spreads good news to periphery: dissemination. This is contrary to policies of putting the country in the driver's seat.

- Knowledge of what works should come from the direct experience of those who are interested, by peer-to-peer benchmarking and horizontal experimentation

- Knowledge brokering, catalyzing the motivational energy existing among the doers.

The current crisis and the challenges for evaluation

The need for a conceptual framework that supports the thrust for autonomy, equity, development.

Improve existing evaluative tools to face new problems:

- Make ordinary tools (rankings, indicators) usable to the people

- Develop alternative designs for impact evaluation

Learning to learn

References

- Argyris C. and D. Schon, 1978, *Organizational Learning: a Theory of Action Perspective*, Reading, Addison Wesley
- Banerjee A. And Duflo E., 2011, *Poor Economics. A Radical Re-thing of the Way to Fight Global Poverty*, Publoc Affairs
- Brinkerhoff R.O., 2003, *The Success Case Method. Find Out Quickly what's Working and what's not*, San Francisco, Berret-Koeler Publishers
- Cartwright N. and Hardie J., 2012 , *Evidence-based policy: doing it better. A practical guide to predicting if a policy will work for you*, Oxford, Oxford University Press
- Dart J. and R. Davies, 2003, "A Dialogical, Story-Based Evaluation tool: the Most Significant change Technique", *American Journal of Evaluation*, 24 (2), 137-145
- Ellerman, D., 2006, *Helping People Help Themselves*, Ann Arbor, University of Michigan Press
- Davidson, J., 2005, *Evaluation Methodology Basics. The Nuts and Bolts of Sound Evaluation*, Sage
- Lederman, S., 2012, "Exploring the necessary conditions for evaluation use in program change", *American Journal of Evaluation*, 33 (2): 159-178
- Patton M. , 2011, *Developmental Evaluation*, New York, the Guilford Press

References, cont.

- Pawson R. and Tilley N., 1997, *Realistic Evaluation*, London, Sage

- Perrin B., 2002, "How to – and How Not to – Evaluate Innovation" in *Evaluation*, vol. 8 (1).

- Preskill H. and T. T. Catsambas, 2006, *Reframing Evaluation through Appreciative Inquiry*, Thounsand Oaks, Sage

- Sabel, C. F. (2004), "Beyond Principal-Agent Governance: Experimentalist Organizations, Learning and Accountability", draft discussion paper prepared for WRR meeting, Amsterdam (May 10-14, 2004), in Ewald Engelen & Monika Sie Dhian Ho (eds.), *De Staat van de Democratie. Democratie voorbij de Staat*. WRR Verkenning 3, Amsterdam, Amsterdam University Press, pp. 173-195.

- Stame N., 2010, "What Doesn't Work? Three Failures, Many Answers", in *Evaluation*, vol. 16(4), pp. 371-387

- Sternin J. and R. Choo, 2000, *The power of positive deviance*, Cambridge, Harvard business

- Stern E. et al, 2012, *Broadening the range of designs and methods for impact evaluation. Report of a study commissioned by the DFIF*, DFID Working Paper n. 38

- Woolcock M., 2009, *Toward a Plurality of Methods in Project Evaluation: a contextualized approach to understand impact trajectories and efficiency,* Working paper 73, University of Manchester: Brooks World Poverty Institute