

## FASTER, SMARTER AND CHEAPER: HYPOTHESIS-TESTING IN POLICY AND PROGRAM EVALUATION

Mark Matthews and Geoff White

We were commissioned by a State government in Australia to review its approach to the evaluation of a major investment in science and innovation that it made over a decade ago. It quickly became clear to us that the traditional auditing process adopted by our client which relied on reconstruction of the original objectives and targets of the investment was inappropriate. It boosted costs, slowed down the evaluation and sapped the goodwill of stakeholders – without contributing much by way of evidence on the impact and value for money of the investment.

Apart from the obvious finding that an evaluation framework should have been specified at the outset of the intervention, we concluded that the audit approach was cumbersome and expensive. It required reconsideration. Accordingly, we suggested an alternative designed to increase the speed, effectiveness and efficiency of the evaluation. Our advice might have wider applicability not only in the science and innovation policy field but more generally for the assessment of major investments implemented in uncertain contexts.

The specific solution we proposed drew on the *structured hypothesis testing* techniques used by the US security intelligence community. It involves assessing the investment by formulating and testing succinct propositions against summaries of available evidence in a structured and sequential manner. It reverses the time-line of the audit approach (tracking the intervention as it unfolds over time) by identifying the contribution of the intervention from a formative perspective – how we might move forward taking account of current budgetary and policy priorities.

The approach aims to maximize the signal to noise ratio by quickly testing hypotheses against the balance of probabilities and by reporting the results in a concise fashion ac-

ording to a format suitable for use at later dates. A critical aspect of the approach is that it allows for, and encourages, an iterative process by which hypotheses are readily reformulated and evaluation results revised as fresh evidence becomes available – without recourse to major and costly re-writing of unwieldy reports.

The hypothesis testing approach we piloted proved to be effective in getting the State government evaluation back on track and engaging stakeholders in a positive dialogue about preparing for the future (rather than just reflecting on the past). It also demonstrated potentially significant cost savings compared with the audit-based approach. Consequently, the State government is now adopting the approach more widely and further developing its key components, namely:

- Engagement of stakeholders in jointly specifying and agreeing the key hypotheses underpinning the intervention program to be tested and adjusted as development proceeds
- Agreement amongst stakeholders at the outset of the program on the design of the monitoring and evaluation questions, framework, methods and reporting arrangements and their respective roles in providing the necessary data
- Mixed evaluation methods and a process of triangulation to assess convergence of the emerging evidence – to reduce complexity as the intervention unfolds
- Tracking mechanisms to ensure that the evaluation can help reduce uncertainty and inform decisions in a clear and simple way especially at anticipated ‘forks’ in the road of program development
- Methods by which evaluation design and reporting can be standardized and the integrity of the evaluation process protected – including protocols that guarantee evaluators’ independence regardless of whether they are internally or externally commissioned.

The approach we piloted is promising for more widespread use given its following compelling attributes.

Firstly, by reducing the complexity and workload of evaluations it lowers their cost and duration. If applied in a continuous developmental manner, such evaluations could spot early program failures and contribute to better use of resources.

Secondly, it offers a clear and standardized discipline for the design and conduct of program evaluations. It enables public authorities to become smarter commissioners of evaluations, recipients of government funding to be clearer about evaluation information needs, and evaluators to be more consistent in the conduct and reporting of evaluations.

Finally, the approach can be set up at the start of a program and at low cost. Hence, program managers will know at the outset what the key evaluation issues are and how they will be assessed. The resulting discipline of continuously monitoring and evaluating how public value is being generated (or not) by testing hypotheses against available evidence increases the likelihood of generating benefits for taxpayers by fostering continuous improvement.

On the basis of experience to date, the hypothesis testing method could help to bridge theory of change approaches and real time developmental evaluation in a pragmatic and rigorous way. It could contribute to the integration of policy and program design with the evaluation methods to be used before, during and after the intervention.

This potential exists because policy and program design in complex and uncertain contexts is itself often based on hypotheses about the underlying theory of change and the appropriate form for program architectures and delivery mechanisms. Consequently, the approach can also be used to collate and analyse the evidence that drives policy

and program design by integrating, assessing and reporting research findings from different interventions in a single comprehensive framework.

The next phase in the development of the approach is a set of pilot activities to be carried out in partnership with interested

Australian federal government departments and agencies to explore the potential of this integrated approach (including pilots in analysing complex evidence that informs policy stances). There is the potential for wider experimentation with developing the approach. For example, it might make sense to set up an 'open source' forum – allowing the re-

sults of multiple pilots to be logged, lessons learned to be spelled out and a community of good practice to be developed. Further information on the approach can be obtained from the authors on request.

The corresponding author is Mark Matthews (mark.matthews@anu.edu.au).

## THE AUTHORS

### **María Bustelo**

Maria Bustelo, PhD in Political Science, is Associate Professor in the Department of Political Science and Public Administration at the Complutense University, Madrid (UCM), Spain. She has worked as an independent consultant for Spanish public administrations and agencies, the European Commission, and NGOs in the field of programme evaluation, and in the areas of community development, health promotion, drug dependency prevention, and gender policies. She is also the director of the Master on Evaluation of Programmes and Public Policies (UCM), and member of the Board of Director Committee of the National Agency for the Evaluation of Public Policies and the Quality of Public Services. She leads the European QUING project team at the UCM (QUING: Quality in Gender+ Equality Policies -2006–2011- www.quing.eu; a research project supported by the VI European Framework Programme). She has a number of publications on analysis and evaluation of public policies and gender equality policies.



### **Mattia Prayer Galletti**

Mr. Mattia PRAYER GALLETTI, a national of Italy, joined IFAD in 1988. Since 1989 he has been working as Country Programme Manager in the Asia and the Pacific Division, responsible for IFAD-supported programs in Bhutan, Laos, Mongolia, Papua New Guinea, China, Vietnam, Indonesia, Iran, the Pacific region and India. In 2011 he moved as Senior Evaluation Officer in the Independent Office of Evaluation. Mr. Prayer Galletti holds a B.A.



in Political Science from the University of Bologna, and two Masters in Economics of Agro-Food Systems (Cremona, Italy) and in Development Studies (Geneva, Switzerland).

### **Margareta de Goys**

Margareta de Goys is the Director of UNIDO's Evaluation Group. She was the Vice-Chair of the United Nations Evaluation Group (UNEG) in 2011/12. Previously she was a partner and senior consultant of SPM Consultants, Stockholm, Sweden. She has been involved in a large number of bilateral and multilateral evaluations, both in the capacity of evaluation manager and evaluator. She was the Co-Chair of the NONIE 2012 meeting.



### **Ann M. Doucette, Ph.D.,**

is the Director of The Evaluators' Institute (TEI) and the Midge Smith Center for Evaluation Effectiveness, and Research Professor of Evaluation and Health Policy at The George Washington University, USA. Her expertise includes the development of rigorous and credible performance measurement systems that target accountability, quality monitoring, outcomes and impact for system, organization, and individual levels of program intervention. This work includes specialized emphasis on measurement, which Dr. Doucette considers fundamentally critical in performance monitoring, evaluation practice, and outcome/impact assessment. Her work in these areas most often emanates from a complex adaptive systems perspective.



### **Dr. Irene Guijt**

is an independent consultant, part-time Research Associate at the Overseas Development Institute and a Visiting scholar at the Australian National University. She has worked in international development for over 20 years, emphasising participatory processes. She undertakes research, system design, facilitation and advisory work on learning-oriented knowledge processes in international development. Recent work includes piloting SenseMaker for evaluation in East Africa and Asia, a programme of work on impact evaluation with AusAID, and piloting participatory impact assessment for IFAD. She is active in global capacity building through BetterEvaluation and in the politics of evaluation as co-convenor of the Big Push Forward.



### **Mark Matthews**

is the Executive Director of the HC Coombs Policy Forum at the Crawford School of Public Policy, Australian National University (ANU). The Forum is a joint Australian Government – ANU policy think tank that focuses on experimental and exploratory policy work carried out at the interface between government and academia. Mark's own research and policy engagement work focuses on governments' role as uncertainty and risk manager of last resort and the implications of this distinctive role for policy formulation, program evaluation and review, and the specific implications for science and innovation policy (including public sector innovation). In addition to a career as a management consultant, Mark has held academic positions in the UK universities of Sussex, Bath and Warwick.

