

Evaluation of Large-Scale Research & Development Programs: Comparison of Evaluations of “Our Rural Landscape” and “Transdisciplinary Tobacco Use Research Centre”

Martin Andrew, Julie McGeary

Martin Andrew, URS Australia, martin_andrew@urscorp.com

Julie McGeary, Victorian Department of Primary Industries (DPI), Julie.McGeary@dpi.vic.gov.au

Poster paper presented to Australasian Evaluation Society’s International Conference, Perth, 10-12 September, 2008

Introduction

There has been a trend to investments in large-scale research and development initiatives, i.e. large in terms of scope of R&D, budget, staffing, disciplines included, and in response to address big challenges in a more holistic manner than can be addressed by individual research grants. Examples in Australia with which we are familiar include the Cooperative Research Centres Program, and programs funded cooperatively by the Rural R&D Corporations such as “Sustainable Grazing Systems” (SGS; Lodge *et al.* 2003) and “Grain & Graze” (G&G), and “Our Rural Landscape” (ORL, an initiative by the Victorian Government). There are many more. But this is a global phenomenon, and Trochim *et al.* (2008) cite similar trends in the USA and Europe in medical research: “... trends suggest that science is getting bigger, in amount of funding per initiative and in numbers of scientists collaborating, and that big science accounts for an increasingly large proportion of total research expenditures”.

In parallel with this trend to bigger research investments, is an increasing demand for evaluation of these investments for greater accountability, to learn how to undertake the research and its organisation better, and to demonstrate the benefits and impacts that have resulted, across the triple bottom line (ie the ‘returns on the investment’). Trochim *et al.* (2008) note that this trend poses “... new challenges for the management of science and particularly for how to evaluate the processes and effects of these considerable investments in research”.

In Australia, at least in the agricultural and natural resources management fields, the practice of program evaluation is well-established. Thus, e.g., SGS was evaluated as a whole at mid-term and again near the end of the Program (Lodge *et al.* 2003), as was the subsequent dryland mixed farming R&D program ‘Grain & Graze’.

Trochim *et al.* (2008) report that in the medical field in the USA, the methodology is poorly developed for evaluating large research initiatives, and that “there was little program theory available to guide an evaluation of this type”. In response to this the US National Cancer Institute has initiated the “Evaluation of Large Initiatives” (ELI) project. This has “... involved looking at what we currently know about evaluating large research initiatives, examining potential evaluation approaches and methodologies, and assessing the challenges and issues that need to be addressed” to identify good approaches to the evaluation of such large initiatives. The evaluation of the “Transdisciplinary Tobacco Use Research Centre” (TTURC), was the pilot study of the ELI. Thus the learnings from this pilot, which Trochim *et al.* (2008) present, serve as a good benchmark for examining evaluation practice.

In 2007 we developed, carried out and reported the end of program evaluation of ORL (McGeary *et al.* 2007). How does the methodology we used for ORL compare with the practice and findings of Trochim *et al.* (2008)?

About the Programs

Transdisciplinary Tobacco Use Research Centre

Trochim *et al.* (2008) report that "... the TTURC initiative is a 5-year \$70 million initiative that provides support to multiple research centers to study new ways of combating tobacco use and nicotine addiction, and to help translate the results and implications of this work for policy makers, practitioners, and the public. Each center's research portfolio covers basic and applied research as well as research on policy-relevant issues in studies being conducted at the center. One of the primary goals of the initiative is to encourage and support transdisciplinary ... research (i.e., research that crosses and integrates theories and methods from different disciplines). Research supported and generated by the initiative is intended to set a new direction in how tobacco-related research should be conducted. Researcher training is a major component of the initiative and includes new and established investigators with the hope of broadening their scope of expertise within tobacco and across disciplines. Specific funds are provided to the centers to help facilitate the translation of basic and applied research into policy and practice."

Our Rural Landscape

Our Rural Landscape (ORL) was a key Victorian State Government strategic initiative, designed and managed by the Department of Primary Industries (DPI) to develop profitable and sustainable farming systems that addressed community concerns, maintained the natural environment and enhanced Victoria's access to international markets. With a total budget of \$107 million, including co-investment, ORL ran over four years from July 2003 to June 2007, was delivered through 15 major projects and involved about 120 DPI staff and multiple collaborators located across Victoria and beyond.

ORL projects included cutting-edge plant, animal, fish and soil microbial genomics research, nanotechnologies for food sciences and genomics for food bio-security issues. Research was conducted which enabled better linking of public policy to the changing socio-economic profile of rural Victoria, and yet other work aimed to foster a culture of innovation within DPI and community stakeholders. While most of the research tended to be at the fundamental end of the R&D spectrum, some work was further along the 'route to market' than others.

ORL itself was part of the Victorian Government's \$310m Science, Technology & Innovation Initiative Program, Phase 2 (STI-2) aimed at improving science, technology and innovation skills, and benefits of commercial, scientific research, collaboration and science awareness natures. The program was linked to the Victorian Government's *Growing Victoria Together* vision statement.

Similarities and differences between the TTURC and ORL Programs

As Programs, ORL shared many similarities with TTURC (in addition to being large in scale). These include:

- collaborative teams or networks of scientists
- multiple research centers
- encourage and support transdisciplinary research
- broader range of outcomes, including the

- collaboration between institutions
- address complex cutting-edge problems ...
- ... not as amenable to individual research grants
- research portfolio covers basic and applied research as well as research on policy-relevant issues
- study new approaches to research problem solving
- social impact, with ...
- specific funding for these activities
- translate the results and implications of this work for policy makers, practitioners, and the public
- large research budgets (relative to grants to individual scientists)
- products included improved methods and models

Differences included that ORL:

- was a 4-year program; TTURC is of longer duration
- had more focus on needs of end-users of the research in ORL; hence ...
- outputs other than just scientific publication were valued and rewarded.

The respective program evaluations

There were strong similarities between the approach to program evaluation in TTURC and ORL, and also some key differences.

Similarities

Similarities included:

- Based on an outcome logic model that depicts the hypothesized sequential causal relationships among outcome constructs. In ORL this was the 'Outcomes Hierarchy', part of the 'input-output-outcome' Program Logic; in TTURC, this was considerably elaborated
- Evaluation began with the development of an evaluation framework for data collection and analysis based on the outcome logic model, including questions to pursue, Performance Indicators and the means of measuring them.
- This evaluation framework was developed collaboratively. For TTURC there was active participation by TTURC investigators, funders, and other stakeholders; in ORL the framework was developed by the external URS evaluation team in liaison with the multi-stakeholder ORL Evaluation Steering Committee
- A focus on short-term and intermediate-term outcomes. (Insufficient time had elapsed for long-term outcomes to be manifest)
- Evaluation included examination of R&D processes and implementation as well as ...
- ... management and organisational structure
- Use of a mixed-methods approach: multiple qualitative and quantitative measures and analyses
- Incorporated judgments of multiple peer evaluators using their written assessments and their ratings of outcomes on quantitative scales
- Participatory approach – input sought from a variety of stakeholders including all of the researchers, research centre staff, and a number of independent peer reviewers

- Evaluation relied wherever possible on pre-existing data sources including standard reports
- New data obtained from stakeholders via bespoke questionnaires to elicit their opinions and evaluative assessments regarding the entire range of outcomes in the logic model
- Evaluation analyses and report were structured according to the evaluation framework.

Differences

There were some key differences also:

- ORL was a one-off, end-of-Program evaluation; whereas the TTURC evaluation is conducted annually and is ongoing (Trochim *et al.* 2008 report the first three years of results).
- By contrast, the ORL evaluation was built into the ORL Program plan and budget; but the TTURC evaluation required special additional funding.
- The ORL evaluators were external contractors; the TTURC evaluators are internal.
- In TTURC the participation in the development of the evaluation plan was much broader than in ORL, and included all of the researchers, centre staff, and a number of independent peer reviewers; in ORL the participation was restricted to the ORL Evaluation Steering Committee members comprising senior managers from the Victorian Dept of Primary Industries and the Victorian Government funders, and a few key research managers.
- In ORL the outcome logic was explicit developed at the outset of the Program and was fundamental to the design of the Program; in TTURC the outcome logic was deduced by analysis of the way the Program had developed. Thus, the TTURC the evaluators developed a "... comprehensive map of the outcome domains that needed to be addressed in the evaluation. brainstormed 262 potential outcomes that were edited and condensed into 97 final outcome statements in 13 clusters and five general regions (Collaboration, Scientific Integration, Professional Validation, Communication, and Health Impacts - essentially clusters of clusters that illuminate a higher level of generality) via multidimensional scaling and agglomerative hierarchical cluster analysis, and average ratings were computed for each statement and cluster of statements. An outcome logic model was developed by arranging the clusters of the concept map in the expected temporal order".
- TTURC's evaluation focused on the production of quality publications and new research methods; ORL's evaluation was broader, reflecting the wider range of outcomes and products that were valued
- Accordingly, TTURC questioned only the researchers, whereas ORL consulted a wider range of stakeholders including end-users
- The questionnaire development process in TTURC was much more analytical and extensive: the questionnaire "went through multiple cycles of structured review and revision with a variety of groups. Psychometric methods were used to assess the measurement properties of four scales that were developed". In ORL the questionnaire was road-tested by a small group of researchers and ORL management.
- TTURC achieved a questionnaire response rate of 92% (216 of 234 researchers); in ORL it was less than 50%. In TTURC the questionnaire was an annual survey and institutionalised as a management expectation – the response rate was made competitive between the different institutes without jeopardising the confidentiality of the responses. In ORL the

evaluation was viewed by many researchers as an additional impost at the crucial time of analysing data and wrapping up their research.

- In TTURC, 26 separate index variables were constructed for analysis from different combinations of question items, with each scale and index score linked to an outcome area on the logic model. In ORL, evaluation power was obtained via carefully constructed Goal Attainment Scaling questions.
- TTURC evaluation incorporated objective content analysis of progress report summaries which were tested for presence/absence of 14 markers.
- Likewise, the TTURC evaluation incorporated an objective bibliometric analysis of the TTURC publications
- In TTURC, the impact of the research was assessed via peer review from other researchers, whereas in ORL it was assessed by consulting the next- and end-users directly, as far as was possible.

Discussion

Trochim *et al.* (2008) made the following ten conclusions about the evaluation of large-scale research initiatives. How well did the ORL evaluation reflect them?

1. Develop a comprehensive conceptual model – *ORL did this*
2. Use participatory and collaborative evaluation approaches – *ORL did this*
3. Incorporate integrative mixed methods – *ORL did this*
4. Integrate evaluation with existing reporting systems – *ORL did this where possible*
5. Adapt the evaluation to the initiative's stage of development – *whilst the ORL evaluation framework elaborated the long-term outcomes, evaluation data could be obtained only on short and some intermediate-term outcomes (which were leading indicators of the long-term outcomes)*
6. Develop standardized cross-initiative evaluation systems – *the ORL evaluation was a bespoke design, but drew from common Australian evaluation practice*
7. Utilize peer review approaches – *ORL did not do this in a formal sense. The focus was on the ultimate outcomes of the research for end-users, rather than the quality of the science as judged by fellow researchers.*
8. Address issues of causation and control – *ORL did this via GAS questions and judgement of informed stakeholders, rather than by formal statistical analysis / 'experimental design' approach of TTURC.*
9. Improve funding and organizational capacity for evaluation. (They recommend an automatic allocation of 1% of budget for evaluation), *and*
10. Address management issues in large initiative evaluation – *the commissioning and management of the ORL evaluation reflected the well-developed and institutionalised approach to evaluation within the Victorian Dept of Primary Industries. The budget for the external evaluator's fees was 0.4% of the DPI's ORL budget; however in-kind costs by DPI, and the cost of other evaluation activities undertaken in ORL additional to this end-of-Program evaluation, would bring the allocation nearer to 1% of budget.*

There were more similarities than differences between the evaluations for TTURC and ORL, which is not surprising given the realities of conducting program evaluations. There were some key differences, however, which largely reflect the different cultures

of the two Programs, with TTURC relying on statistical approaches to the development of its design and assessment of the quality of its products (research reports and papers) whilst ORL was more focused on achievement of longer-term outcomes and the assessment of that judged by a range of informed stakeholders.

References

Lodge GM, Mason WK, Andrew MH, Anderson CA and Muir LL (eds) (2003) *Special Edition for the Sustainable Grazing Systems Program. Australian Journal of Experimental Agriculture* **43**(7-8): 663-1082

McGeary, J, Andrew, M, Smith, S, Burnside, D, Pechey, L, Ritchie, T (2007) Managing a complex evaluation – the case of “Our Rural Landscape”. Presented at the Australasian Evaluation Society’s International Conference, Melbourne, 3-7 September, 2007

Trochim, WM, Marcus, SE, Mâsse, LC, Moser, RP and Weld, PC. (2008) The Evaluation of Large Research Initiatives: A Participatory Integrative Mixed-Methods Approach. *American Journal of Evaluation* **29**; 8-28