A Program Logic Approach to Evaluating Educational Innovations

Prof. Caroline Crosthwaite,
University of Queensland, Brisbane, Australia
c.crosthwaite@uq.edu.au

Dr Lesley Jolly,
University of Queensland, Brisbane, Australia
ljolly@bigpond.net.au

Lizzie Brown,
Engineers Without Borders, Brisbane, Australia
l.brown@ewb.org.au

Abstract*: This article argues that good program evaluation needs to treat educational innovations and processes as it would any other social program. Drawing on the literature in social program evaluation, we argue for the need to understand the theory behind the project which embodies assumptions about how its effects are produced. Once these are articulated and examined, they can then be incorporated into a program logic which lays out inputs, activities, outputs and outcomes and can be used as a planning document for future curriculum development as well as a template for evaluating the program. Some preliminary data from a current evaluation of the use of sustainable development projects sponsored by Engineers Without Borders Australia exemplify how such an evaluation produces fine grained understandings of the mechanisms at work and their effectiveness. The challenge is thrown out for all engineering academics with an interest in education to incorporate these evaluation approaches into their practice.*

Context and research questions

The EWB Challenge is a national design program for first-year university students coordinated by Engineers Without Borders Australia. The program aims to enhance students' first-year learning experience and develop their core competencies through team-based design focused on real and inspiring sustainable development projects. In 2008 and 2009 the in-country partner organizations were both working in Cambodia. At the University of Queensland over 950 freshmen completed the projects in each year as part of a course called ENGG1000: *Introduction to Professional Engineering*. The course was split between a lecture component concentrating on professional skills and group work sessions in which small groups of students worked together to come up with a sustainable design solution to a local problem. Students from all divisions of engineering, including software, mechanical, chemical, civil and mining engineering had projects to work on centred on a Cambodian setting. Although the students did not have the opportunity to visit Cambodia or interact directly with the clients for the projects, in 2008 82% of students said that it was important to them that the projects were real-life ones. In addition 56% of students said that the experience broadened their understanding of what engineers do and enhanced their university experience. There therefore seemed to be some benefits of the projects which could be built upon further in the first year and beyond. But there were also some difficulties with the course as a whole including inadequate integration of the projects with other components of the course. Some redesign was felt to be necessary but for optimum results we needed to know exactly what impact the course and its various components was having and how.

A standard way of evaluating the impact of a course of study on the students receiving it is to ask them to respond to propositions such as "this course was intellectually stimulating" or "I felt I belonged to a group of students and staff engaged in enquiry and learning". Apart from all the possible external

confounds such as mismatches of teaching and learning styles, student resistance to particular subject matter and so on, this kind of evaluation approach has some internal limitations that make it hard to generalise success or pinpoint and remediate failure (Soundarajan 2004).. Was it subject matter or delivery that the students found stimulating? If delivery, did that depend on the teacher or were there structural elements in play? What was the outcome of the intellectual stimulation, and is that something we were aiming for? And for the students who did not agree with the proposition, where did we fail them? In short, such questions embody a range of assumptions about teaching and learning outcomes which our evaluations rarely tests and hence can rarely build on. To capitalise on the potential of the EWB projects we needed a finer grained analysis of their strengths and weaknesses and we found the necessary tools in the discipline of evaluation research.

Theoretical Framework

Program evaluation as a discipline grew out of social science and is concerned to establish the impact and worth of social interventions, often in areas such as criminology, public health and welfare. Throughout the twentieth century the majority of such evaluations tended to measure outcomes only and there are early examples of those in the field of education research. Those educational evaluations generally looked at whole curricula rather than local interventions (Weiss 1998). As the discipline of evaluation research developed it became clear that treating social interventions like a dose of medicine and investigating the state of the patient's health before and after taking it was too simplistic. These interventions happen in complex social situations in which the circumstances and attitudes of the recipients of the intervention affect the outcome as well as all the other things that are happening to them at the same time as the intervention. Added to that every intervention creates emergent processes of its own involving the politics of choice and compliance. For instance when we introduce a new course or pedagogy or technology to the teaching situation we can measure impacts such as failure and attrition rates relatively easily. We can even set up control groups who are not exposed to the intervention and compare them to the test group. We can do pre-tests and post-tests and all the paraphernalia of experimental and quasi-experimental method to claim that intervention x had impact y. But unless we ask about the nature of the cohort, what the impact of concurrent courses and institutional events was and examine the mechanisms by which the measurable impact was actually achieved, we don't really know that it was the intervention that made the difference. Many of us will be familiar with the situation where an intervention that is successful initially gradually fails subsequent cohorts without our being able to say why. Evaluation approaches which take a strictly experimental approach and ignore all the contextual factors have been attributed (Pawson and Tilley 1997) with this common failing of program evaluation to really explain why an intervention works well in one instance and not in another, or only achieves part of its objectives. Such experimental evaluations are still popular in the field of education (Rudd and Johnson 2008).

Nowadays every textbook on program evaluation insists on the necessity to include contextual factors. For instance one Australian author says it is the task of the evaluator to "understand[] the nature of interventions or evaluands such as:

- Their internal structure and functioning;
- Constraints that shape design and delivery; and
- Societal factors that influence the development of evaluands, how evaluands themselves change over time, and how in turn the evaluand contributes to social change" (Owen 2006).

The internal structure and functioning of a program refers not just to management and implementation issues but the mechanisms by which they achieve their effects. Some of those mechanisms are often implicit or even invisible (Pawson 2009) but they all need to be exposed before we can fully describe the nature of the intervention. In this area the literature is rather confused thanks to a proliferation of terminology such as "policy theory", "program theory", "program logic" or "log frames", some of which have come from the incorporation of theory-based evaluation into applied areas such as international development.(Rogers 2007). In this article I follow Leeuw (2003) in making a distinction between program theory, which specifies how and intervention causes the intended or observed

outcomes, and program logic which specifies inputs and components as well as outputs and outcomes and the links between these elements (see Figure 1).

Arriving at program theory, then, is one part of developing a program logic. One may or may not wish to or need to evaluate every aspect of the program logic. Turnbull (2002) for instance suggests that validating the causal mechanism/s across different settings may be sufficient. She uses the example of a study of effective management teams in education which included some assumptions about the effect of team cohesion on the outcome. Team Cohesion as a mechanism that has been established as a causal mechanism with particular outcomes in business settings. If Team Cohesion could be shown to be a causal mechanism with similar outcomes in educational settings, it would then be unnecessary to retest the whole model of team effectiveness which included this mechanism.

In our study we need to expose the program theory as completely as possible in order to capture all relevant mechanisms so that we can build on them in developing the course in question and using its benefits in subsequent courses. This is in part a result of the large number of staff involved in teaching a course which has various different components and a very large enrolment. It would be surprising if people were not operating with divergent and perhaps contradictory assumptions and this has the potential to affect outcomes. Once we have done that we will be able to build up a program logic specifying what is needed to reach desired outcomes, such as the example in Figure 1, which will help in curriculum planning.

Inputs	Activities	Outputs	Initial Outcomes	Intermediate Outcomes	Long-term Outcomes
Lincoln/Marathon County UWEX Ag Development Agent/Office resources Lincoln/Marathon County Conservation Staff/Office resources Central Wisconsin River Graziers Farmer Network Marathon County Chamber/Economic Development Corp. staff and resources.	Research, develop, pilot and educate on: • Farmstead/farm business assessment tools. • Business planning tools and training for farm applications. • Farm business transfer alternatives. Develop Mentor Network • Develop mentor training program and support materials.	Number of: • Farmsteads/farm businesses evaluated for determining transferability. • Farmers accessing and/or utilizing financial and business planning/management resources. • Farms aided in the transfer of ownership. Number of : • Existing farmers receiving training on mentoring skills.	Existing and potential farm operators will have: • Increased knowledge on evaluating the transferability of farmsteads/businesses. • Increased knowledge of business plan development. • Increased knowledge of ownership transfer options.	• Existing and potential farmers will evaluate farmsteads/businesses to make objective transfer and purchase decisions. • New farmers will create and implement business plans. • Existing farmers will create and implement ownership transfer plans.	New agribusiness enterprises have increased their ability to stay in business. Existing farmers have improved their ability to profitably retain or transfer their farm businesses.

Figure 1: example of program logic from a agribusiness project (downloaded from http://www.uwex.edu/ces/pdande/evaluation/evallogicmodelexamples.html)

Methodology

In 2004 the National Research Council of the National Academies in the United States published a meta-study of 698 evaluations of mathematics curricula (NRC 2004). Our approach is based on that study's recommendations for the design of a scientifically valid evaluation. The first step is to produce a description of the program theory behind the curriculum innovation which encompasses both content and implementation as shown in Figure 1.

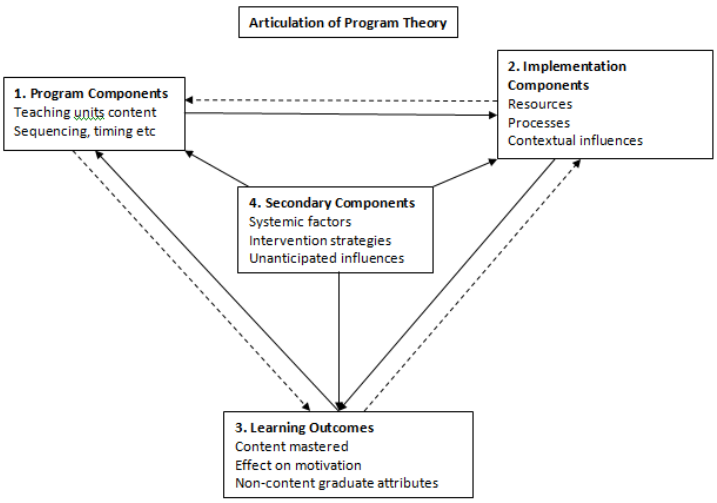

Figure 2: Primary and secondary components of evaluation based on NRC 2004, p.40

Methods used in data collection include content analysis of documentary sources, interviews and focus groups with staff and students and observation of lectures and group work. Documentary sources include course and curriculum documents but also sections of the reflective journals kept by students during the course. The interview protocols are based on the realist approach advocated by Pawson and Tilley (1997) which seeks to articulate participants' theories of what is going on through a process of conceptual refinement. The underlying thrust of this kind of interview is to find out 'what works for whom in what circumstances' and it is based on interviewer and interviewee working towards a shared vocabulary to describe the situation. This combination of methods places our methodology within Leeuw's (2003) "policy-scientific" approach which he suggests is best suited to ex post evaluations of programs and policies backed by documentary evidence and, implicitly, being used to develop and refine further policy. In our case curriculum is the analog of policy.

The course in question is run during the first Australian semester of the year from February to June. Investigations began in September 2008 with analysis of student reflective journals from earlier in that year, focus groups with students and tutors and some staff interviews as well as examination of curricula documents. A new course co-ordinator took over in 2009 and he was concerned to address students' dissatisfaction with the integration between the lecture strand of the course and the project work and his sense that the course lacked an overall coherence. Further interviews have been conducted during first semester 2009 and observations of lectures and group work is currently under way. All text was imported to NVivo8 and analysed using that program. The following section details a part of the preliminary results.

Findings to date

During the course students were required to keep an online reflective journal every week. On one occasion they were asked explicitly to reflect on the EWB Challenge, and offered the following prompt questions:

This is a real-life project, did it matter to you that you were designing solutions for real people in need? In what way did it matter or not matter to you?

How did this project help you to learn about sustainable development?

How has your perception of what an engineer "is" and "does" changed since you began the Challenge?

Do you think the EWB challenge enhanced your university experience in first semester, or would a "made-up" scenario worked just as well? Why?

791 entries were recovered although not all of these are complete and some do not answer the question posed. After first-pass analysis of 350 entries it was decided that categories has been exhausted and we proceeded to incorporate this data with that received from 10 focus groups involving 50 students. These focus groups were held in October nearly a semester after the completion of the projects. For reasons of space we will deal with just two findings here; what students learned about sustainable development and what motivated them to perform well in the projects.

By the time of the interviews students' understanding of sustainability was sketchy at best, reduced in many cases to an ability to remember the phrase "triple bottom line". Analysis of the journals had revealed that even when students could say more than that, the economic aspect of sustainability dominated. Some reason why these projects may produce this result is revealed by comments such as this:

> *Local materials should be used and if possible most of the system should be able to be built by the people there. Economic sustainability has a much higher importance to the poor and so I think the triple bottom line becomes skewed. After all, we're talking about people's lives and I believe that should come before the environment.*

Worryingly, nearly 10% of students claimed sustainability was not relevant to their projects (which were sustainable design projects). A further 10% identified sustainability as a problem of the developing world while only 5 students admitted that it was a problem of their own society too.

As far as motivation went, the outstanding finding was that students were highly motivated by the opportunity to improve the lot of other human beings, they enjoyed being "problem-solvers" and the friendships produced by working together in groups. Such fine grained knowledge of the nature of the cohort and the outcomes can clearly be utilised in refining this course and planning for subsequent ones where the level of knowledge about sustainability and the motivating power of altruism can both be part of the inputs.

Early in the project we articulated part of the program theory in Figure 3. This analysis of what the students told us in two different ways and at different times helps us address some of the questions raised by the draft theory.

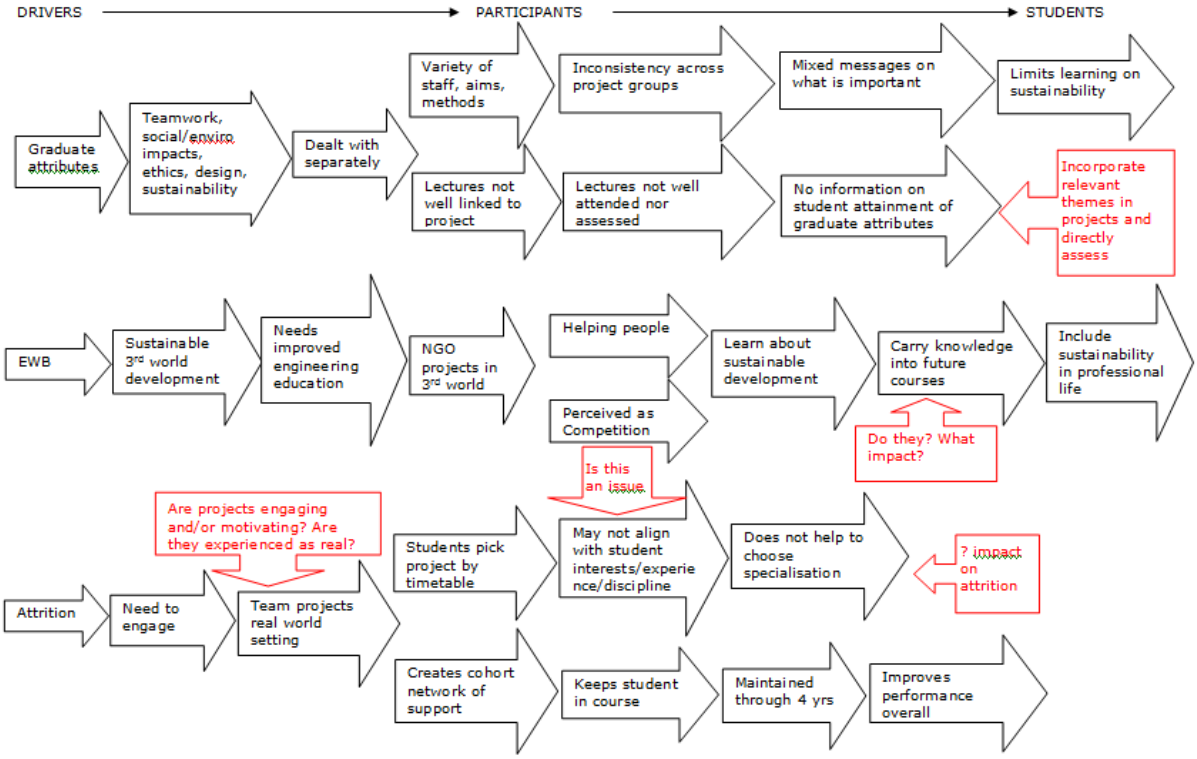

Figure 3: Part of draft program theory for ENGG1000

However, this representation of the program theory does not include some mechanisms recently described by staff during interview. One of the concerns for future curriculum design is the amount of project work to be included in any year. Great claims are sometimes made for projects while some staff feel that it doesn't leave enough time for the basics to be covered. When asked how they thought project work produced any beneficial effects it had, staff started to talk about they way engineers approach practical problems in an empirical "try it and see" way. Projects at this introductory level were thought to be good for introducing students to this approach and thus for introducing them to the way engineers work, an objective of the course. In terms of future articulation of program logic, this allows planners to consider what kind of activities should be used. Whole projects may be only one of several possibilities for utilising and refining an engineering problem solving skill.

Indications for the future

We could get better answers to the question "What works for whom under what circumstances" if we could compare the implementation and outcomes of the EWB projects across a range of settings. Luckily, nearly every university in Australia uses the EWB projects to some extent and at time of

writing a team of academics from seven universities is seeking funding for a comparative evaluation. The different universities site the projects differently within courses and allow them different weight in the assessment. Comparative analysis of this data will allow us to describe relevant dimensions of context (on campus or remote, place in overall curriculum etc) and the mechanisms at work (Problem Based Learning, collaborative group work, experiential learning etc) and the connections between the two which produce observed outcomes. The resulting matrix of Context + Mechanism = Outcome (C/M/O matrix) for a large number of variables will then be available for subsequent curriculum renewal.

But evaluation needs to be an ongoing process and it takes considerable time and resources to do well. External funding cannot always be available but evaluation is always needed. Turnbull's article suggests ways in which faculty could "rationalise" the process to some degree but the need to evaluate continuously (not always the same course or the same component or for the same reason, but fairly constantly) suggests there is a role for regular engineering faculty to build evaluation skills and processes into their everyday practice, perhaps using externally funded projects as the starter and basis for ongoing efforts. For this to happen, staff will have to be supported to acquire the necessary research skills and manage the process. There will have to be recognition of this work as legitimate research activity and commensurate rewards. Now there's a challenge for the engineering education community!

References

Leeuw, F. (2003) Reconstructing program theories: methods available and problems to be solved. *American Journal of Evaluation 24*(1):5-20.

National Research Council (NRC) of the National Academies (2004). *On evaluating Curricular Effectiveness*. Washington, D.C.: National Academies Press.

Owen, J.M. (2006). *Program Evaluation: Forms and Approaches*. Crows Nest, NSW:Allen & Unwin.

Pawson, R. (2009). Invisible Mechanisms. *Evaluation Journal of Australasia 8*(2):4-13

Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London: Sage.

Rogers, P. (2007) Theory-Based Evaluation: Reflection Ten Years On, In S. Mathison (ed.) *Enduring Issues in Evaluation. New Directions in Evaluation No114*.(pp. 63-82). San Francisco, CA: Jossey-Bass.

Rudd, A., Johnson, R.B. (2008). Lessons learned from the use of randomized and quasi-experimental field designs for the evaluation of educational programs. *Studies in Educational Evaluation 34*:180-188.

Soundarajan, N. (2004). Program assessment and program improvement: closing the loop. *Assessment and Evaluation in Higher Education 29*(5): 597-610.

Turnbull, B. (2002). Program theory building: a strategy for deriving cumulative evaluation knowledge. *American Journal of Evaluation 23*:275-290.

Weiss, C. (1998). *Evaluation*. Upper Saddle River, N.J.: Prentice Hall.

Acknowledgements

This research has been supported by a grant from the University of Queensland Teaching and Learning Strategic Grants Scheme and the active participation and support of Engineers Without Borders staff.

Copyright statement