Learning from evaluation in other sectors: adjusting to the changing demands on evaluation and evaluators by joining with mainstream evaluation

Keryn Hassall, ARTD Consultants, Sydney, Australia

Abstract

Increasing demands are being placed on energy policies and programs: to contribute to a range of policy outcomes, and for these programs to be evaluated to inform decision-makers. As we seek to evaluate for non-energy impacts we must look at different outcomes, particularly those currently the focus of evaluation in other sectors such as health and social welfare. As governments prioritise emissions reduction, different audiences are using evaluation findings (different agencies, professions, etc). They have their own contexts and expectations of evaluation. As energy efficiency is becoming seen as a means to other policy objectives, there are different clients for evaluation, with quite different understanding and expectations of evaluation. Evaluation professional societies and some governments are establishing standards for evaluation, based on the practice of evaluation in the health, education and social sectors. These may require a different approach to evaluation. It will be challenging to meet these high and changing expectations for programs and evaluation.

This paper suggests that it is time for the field of energy program evaluation to move away from its explicitly technical focus, and to join with mainstream evaluation – the body of knowledge and practice that is maintained and advanced by the professional evaluation societies internationally. It describes the diversity of mainstream evaluation and the potential for collaborating to develop a comprehensive approach to energy program evaluation.

A personal journey through evaluation

My name is Keryn, and I work as an evaluator in Australia. I am a member of the national professional society for evaluators (the Australasian Evaluation Society) and I work under its code of professional conduct. For about 15 years I have worked on programs to bring about change in our society, such as:

- changing the way doctors treat chronic diseases, as part of hospital demand management
- changing the way police and community services deal with family violence
- changing the way a range of social and health services were funded
- developing community capability to deal with social disadvantage, through 600 community organisations working to meet local needs.

You might wonder what that has to do with IEPPEC. For the last 8 years I have focused on programs to reduce emissions and increase energy efficiency. In this time I have worked as a social scientist with engineers, developing energy efficiency and emissions reduction plans in industry. I have worked to evaluate the national mandatory energy audit scheme, aspects of a white certificate scheme, and currently a large market transformation program.

Observations in the field of energy efficiency programs

As an evaluator, I have often studied failure. I explore why a program is not achieving the expected outcomes, whether through problems in design or implementation. When I moved into working on energy programs, I saw people being challenged by problems that I had seen resolved in other sectors. These challenges were not essential to energy programs but were due to the energy professionals being unaware of the knowledge and experience available from other sectors. These

other sectors had built capability over a longer history of delivering programs, and through undertaking detailed evaluation to understand programs in their real-world context.

The growing focus on multiple benefits brings energy program evaluation into the domain of mainstream program evaluation. I'm here to explain why this is a great opportunity to learn and get better information from evaluations, through collaboration. By 'mainstream evaluation, I mean the decades of development in evaluation for programs in health, education, and across the social sectors. This is the body of knowledge that is maintained and advanced by the professional evaluation societies and is shared in their journals and courses.

In 2015, IEPPEC and IEA held an evaluators workshop to discuss approaches to evaluating the multiple benefits of energy efficiency, specifically the health and wellbeing benefits. Through this workshop it was apparent that I was the only participant familiar with how health and wellbeing benefits are considered in social program evaluation. I felt sad to see people trying to work out solutions to challenging problems, without involving mainstream practitioners who have the methods, theory and experience that is needed to address these new information needs.

For this IEPPEC conference, I proposed a paper where I would describe major learnings and changes in mainstream evaluation practice, in the hope that this would provide some guidance of 'lessons learned', and detours to avoid. I hoped to talk about method and technique. But in undertaking the research to prepare this paper, I realised that the most important thing is ***how we think about evaluation***, and less about what we do and how we do it (although that is of course important). What is most important (I discovered) is ***the differences that lie behind the differences we see***.

What do we mean when we talk about evaluation?

A common definition among the evaluation profession is that *evaluation is making decisions about the value (merit, worth, or significance) of something*[1] – and in our case our focus is evaluation of programs and policies. In this definition, *merit* refers to the intrinsic value (quality) and *worth* refers to the instrumental value (utility) of the something being evaluated.

But that isn't enough to tell us what we're talking about when we talk about evaluation; we commonly use the word *evaluation* in various ways. Evaluation as a product (the report), evaluation as a process, evaluation as a profession or practice, and sometimes evaluation as a field of scholarship. Today when I talk about evaluation I'm talking about the practice of evaluation and linking to ideas from evaluation scholars.

The technical approach to energy program evaluation is not very helpful

As a person who works both in mainstream evaluation and energy efficiency programs, I see significant risks from keeping the practice of energy program evaluation separate from mainstream evaluation.

Mainstream program evaluation has been developing as a field of practice since the 1960s in the area of education, health and social programs. It has faced challenges, experimented and made mistakes, learned, and continues to evolve. Without engaging with the mainstream and learning from their experience, energy programs will make many of the same mistakes in program design and evaluation. (Which is inefficient, and if we care about addressing climate change, we do not have time to make many mistakes in our programs). Program evaluation has developed a broad

[1] This definition came from Michael Scriven, an eminent thinker in the evaluation field who has coined a number of the common terms in program evaluation.

range of evaluation-specific theory and practice that has been tried and tested – you can take that learning and build on it. If you are not aware of this body of knowledge you are possibly 'reinventing the wheel' rather than 'standing on the shoulders' of others.

The methods used in energy program evaluation are a narrow subset of the wide range of available evaluation methods. A limited range of methods means a limited range of knowledge can be developed, and only a narrow range of policy-maker questions can be answered. The central role of energy use metrics in energy efficiency evaluation means that people are often left out of the equation. By focusing only on what can be measured or modelled, these evaluations only see part of the picture, and can only tell a partial story about change (Moezzi 2015).

The standard approach to energy program design makes a distinction between technical and social aspects. The energy efficiency gap is explained in terms of barriers – principally the social and psychological aspects that mean the economically and technically efficient ideal is not realised. (Shove 1998). So it appears that humans are at fault for not behaving as predicted by the engineers and economists who estimate the energy efficiency gap. Maybe another explanation is that engineers and economists do not have sufficient understanding of humans, and could reconsider their approach to better reflect human reality.

Much of the practice of energy program evaluation was developed in the USA, to fit the context of state energy policies and to meet the specific needs of their energy program delivery agencies. Energy efficiency programs were developed for 'resource acquisition' – to implement projects which created measurable energy savings in a cost effective way – that could be considered a source of supply for the electricity sector. For these programs, the goal of evaluation was to quantify the amount of energy saved and the cost of achieving this (Blumstein et al. 2000, Lutzenhiser 2014).

The context for energy efficiency programs is quite different in most other countries. People who commission evaluation do not have the same goals for their evaluation as in US program evaluation; they are not seeking the same knowledge and therefore need a different type of evaluation. In a time when decision-makers are seeking greater knowledge from evaluation, and expanding the range of questions of interest to include non-energy impacts, the explicitly technical approach to evaluation is not helpful.

I suggest that it is time for the field of energy program evaluation to reconsider its purpose and break out of its historical focus on resource acquisition, technology, transactions, and verification. By engaging collaboratively with mainstream evaluation, energy program managers can have access to a much greater range of evaluation approaches to provide the knowledge you need and inform future program design.

Learning from mainstream evaluation

From my exploration of the literature, discussions with evaluators and personal reflection, I have selected seven attributes of mainstream evaluation that I think are most important for IEPPEC participants to know about.

- evaluation is a broad practice
- evaluation answers evaluative questions
- evaluation theory is an important basis for practice
- evaluation makes judgements about value
- evaluation uses program theory
- evaluation is interested in causality
- evaluation is sensitive to the context of the program and of the evaluation

I hope that by knowing more about mainstream evaluation, you will be keen to engage with the mainstream evaluation practitioners in your region and find ways to collaborate to build a more comprehensive form of energy program evaluation.

Evaluation is a broad and pluralist practice

There are many varieties of program evaluation. It has a number of roles across the program cycle, and can be used to meet a range of purposes. Evaluation is also a pluralist practice, with a diversity of perspectives and approaches reflecting the different philosophical strains within social science, and the different contexts in which evaluation is used (Schwandt 2015b). One helpful way to look at the breadth of evaluation is the 'five forms' categorisation of evaluative enquiry (Owen 2006). This shows the role for evaluation at all stages of the program cycle. Each of these five forms is used for different purposes. There is also a wide range of evaluation approaches within each of these forms, reflecting the pluralism within evaluation[2].

Proactive evaluation - before the program - to inform the development of a program.

Approaches include: needs assessment, research review, review of best practice.

Clarificative evaluation - early stage of a program - helps to make an intervention explicit by clarifying the program theory (how the program is designed to achieve its outcomes)

Approaches include: evaluability assessment, program theory development/documentation, implementation logic & planning

Interactive evaluation - during delivery - to provide knowledge for decisions about delivering and improving the program

Approaches include: 'developmental evaluation', 'empowerment evaluation', 'responsive evaluation', participatory evaluation, continuous improvement, action research, quality review

Monitoring evaluation - ongoing, for a well-established program - focused on informing program management about implementation and progress towards outcomes.

Approaches include: component analysis, devolved performance assessment, systems-level analysis, use of performance indicators (qualitative or quantitative)

Impact evaluation - to assess the impact of a settled program, in terms of the criteria selected to judge its merits and worth.

Approaches include: process-outcome studies, objectives based evaluation, realist evaluation, goal free evaluation, needs-based evaluation, performance audit.

These evaluation forms and approaches are used in different situations to meet different evaluation needs. The form of an evaluation depends on the stage of the program and needs of decision-makers. The choice of evaluation approach is more flexible, it depends on the context (of the program and evaluation) as well as the skills and background of the evaluator.

The pluralism in evaluation may look daunting for people new to mainstream evaluation – there is no 'one right way' or 'best practice'. A useful way to look at pluralism is the tree of evaluation theorists (Alkin 2013, and earlier editions). It sets out the roots of evaluation and three branches – use, methods and valuing – showing the main evaluation theorists in each branch.

[2] These approaches are listed in Owen (2006) and there are many other approaches available in each of these forms. There is considerable literature available on each of these approaches. Some are general terms to describe an approach; others in quotes are specific approaches developed by particular evaluation specialists.

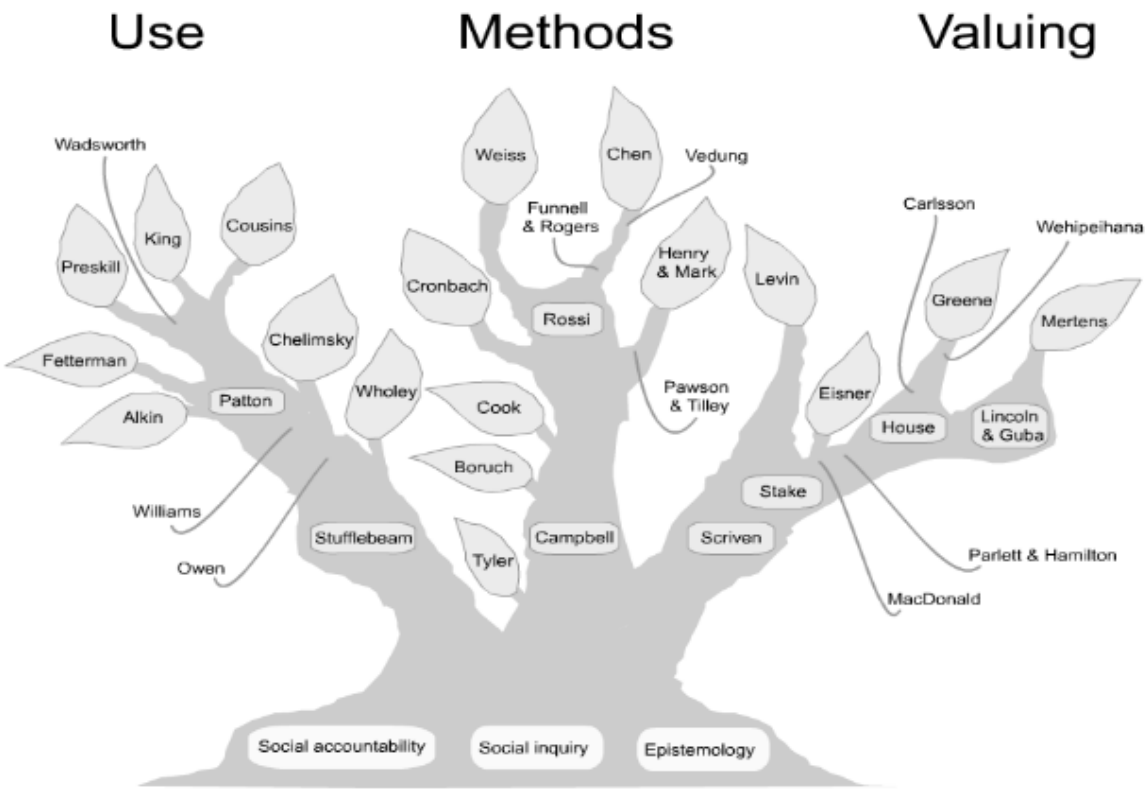

Figure 1: Evaluation theory tree (Alkin 2013)

The pluralism within evaluation is also shaped by the diversity of people who come to evaluation from other fields, and the information needs they aim to meet. Their choices as practitioners are shaped by their beliefs, backgrounds, training, experience, context and the expectations that others have of their work.

This pluralism is a great advantage for the energy program community. Now that governments seek to evaluate the non-energy impacts of programs – such as health, welfare and employment – you have the opportunity to collaborate with the mainstream evaluation profession to answer questions that cannot be answered by standard energy program evaluation.

Evaluation answers evaluative questions

An evaluation is usually designed to answer a few key evaluation questions – those which are important to decision-makers and funders. These questions should be evaluative and may look something like this:

- Was this program worth doing, and it is worth continuing to support?
- What was the quality of the program's design and how well was it implemented?
- How valuable was the project to the participants and broader society?
- What were the factors that made the difference between successful and disappointing implementation and outcomes?
- What else was learned (about the problem and how to address it, mechanisms that cause outcomes, how to deliver it better next time, etc)[3]

Measurement, metrics and indicators are not answers to evaluation questions. Metrics and indicators may be important information for management, and can help answer an evaluation question (Chelimsky 2015). It may be tempting to compile relevant facts and figures, call it "the evidence" and leave it to decision makers to work out the "so what?" But the defining feature of

[3] This list of questions comes from the work of Jane Davidson, a prolific contributor to conferences and online discussions, with a strong focus on practical and methodologically sound evaluation (realevaluation.com).

evaluation is that it is evaluative and has to answer these important questions. Good evaluators learn to take the range of information and data – multiple lines of evidence – and use this evidence in combination to form evaluative judgements in order to answer the evaluation questions.

Too often evaluations are built around specific methods (techniques the team is familiar with) and not around meaningful evaluation questions. This has been a problem both in mainstream evaluation and energy program evaluation. Now mainstream evaluation leaders and academics are advocating better quality evaluation questions, and less focus of evaluation on 'doing' the methods.

Answering evaluation questions requires a range of methods and techniques – using both standard research methods (like surveys, data analysis, document synthesis, or interviews) as well as evaluation-specific methods. These evaluation-specific methods include methods to synthesise information for evaluative judgement, establish evaluative criteria and find a balance across criteria (Davidson 2005, Chelimsky 2015). The use of these evaluation-specific methods enables mainstream evaluation to provide meaningful answers to evaluation questions, supported by the use of evaluation theory.

Evaluation theory is an important basis for practice

Evaluation theory is what makes evaluation a distinct profession. It has its own body of professional knowledge. There is not a single theory but a range which have developed from different evaluation approaches in different contexts.

There has been a tendency for evaluators to focus on method and technique, but evaluation leaders keep reminding us of the value of theory.

> *…if you do not know much about evaluation theory, you are not an evaluator. You may be a great methodologist, a wonderful philosopher, or a very effective program manager. But you are not an evaluator. To be an evaluator, you need to know that knowledge base that makes the field unique. That unique knowledge base is evaluation theory.* (Shadish 1998)

We know that experience alone in a job does not guarantee learning. It is the time spent reflecting, considering and interpreting experience that helps a person to learn and grow in their profession (Schwandt 2015b). Evaluation is primarily a practice-driven field, but it is much more than a simple collection of methods and techniques that can be used by anyone. Carol Weiss, one of the early shapers of evaluation, reminds us that "there is nothing as practical as good theory".

Evaluation theory provides a basis for good practice because it is a coherent body of knowledge which builds on social science theory with additional knowledge about programs, service delivery, effectiveness and evaluation-specific methodology. Without understanding evaluation theory, there is a risk of just 'doing' the methods without seeing how they fit together (or don't) to enable evaluation.

> *Learning the latest methodological advance—whether it's some new statistical adjustment for selection bias or the most recent technique to facilitate stakeholder dialogue—without knowing the relevant theory is a bit like learning what to do without knowing why or when.* (Mark 2005)

Evaluation theory is diverse, and there is much to read if you want to know more. But it need not be intimidating; there are only five essential areas of evaluation theory (Shadish, Cook & Leviton (1991), relating to:

- Knowledge: understanding what is credible knowledge and how this can be produced
- Use: how to use knowledge and information about programs, and produce knowledge that is useful
- Valuing: the role that values and valuing play in evaluation, how to construct value judgements and make values explicit

- Programs: how programs are designed and work to bring about change and how programs can be improved
- Practice: how evaluators should practice in 'real-world' settings.

If you see yourself as an evaluator, are you aware of these five areas of evaluation theory? How do you remain engaged with theory and how does it inform your work and continuous improvement? If you commission evaluations, are you considering whether your contractors are technicians (undertaking methods and techniques) or whether they are professional evaluators who practice with an understanding of evaluation theory.

The importance of evaluation theory is one of the reasons why I'm not encouraging you to learn techniques from mainstream evaluation to bring into energy program evaluation. Instead I want you to be aware that mainstream evaluation has much to offer the energy program community. All you need to do is to bring your energy program expertise, and work collaboratively with the mainstream evaluation community to create a mainstream approach for energy program evaluation.

Evaluation makes judgements about value

> *The distinguishing feature of evaluation as a practice is that it directly engages questions of the value of programs and policies. Evaluation is a judgment-oriented practice—it does not aim simply to describe some state of affairs but to offer a considered and reasoned judgment about the value of that state of affairs.* (Schwandt 2015b)

Program evaluation is a systematic process to make decisions about the value (merit, worth, or significance) of a program or policy. These decisions about the value are based on relevant criteria; for example defining what 'poor' or 'better' means in the context of that program. Underpinning these criteria for judgement are values. These values may relate to the purpose of evaluation, what constitutes worthy information, how information is perceived, as well as values relating to the program and its outcomes and the problem being addressed (Yarbrough et al 2011)

Identifying and being clear about the values shaping an evaluation is considered good practice in mainstream evaluation. The US Standards for Program Evaluation (Yarbrough et al 2011) states: *Evaluations should clarify and specify the individual and cultural values underpinning the evaluation purposes, processes and judgements.* These standards were developed originally for the education sector, and are now widely accepted as good guidance for professional evaluation practice.

Energy program evaluation in the USA is often focused on cost effectiveness for energy saved, as this is a key evaluation question for US decision-makers in their regulatory context. The values implicit in these evaluations relate to rationality and efficiency – both technical and economic.

However other countries have a different rationale for energy efficiency policy and programs, based on different values. These programs need an approach to evaluation that reflects their own values.

For example, programs to address fuel poverty or the health impacts of cold homes. Evaluation of these programs would seek to understand the relationship between energy costs, housing quality, access to health services, wellbeing, disease symptoms and death. In situations with dire health consequences, evaluations may also look at the costs of healthcare and early death. The choice of evaluation approach will depend on values. What do we as a society value more, wellbeing or cost-saving? An evaluation focused on costs to government may conclude that it is better (more cost effective) to let elderly people die in uninsulated homes than to invest in upgrading homes and providing greater access to health services. In a modern democratic society we generally value wellbeing for all people. Relying on the technical process of economic valuation would draw on values that are not consistent with the rationale for the policy or program.

Evaluation uses program theory

> *A program theory is an explicit theory or model of how an intervention contributes to a set of specific outcomes through a series of intermediate results. The theory needs to include an explanation of how the program's activities contribute to the results – not simply a list of activities followed by the results, with no explanation of how these are linked apart from a mysterious arrow* (Funnell and Rogers 2011).

One of the more visible evolutions in mainstream program evaluation has been the increased use of program theory to explain how a program will cause outcomes. Program theory is more than a simple logic diagram or logframe table.

Program theory has two key elements[4]. The ***theory of change*** sets out the causal mechanisms by which change comes about (for an individual, organisation or community). The ***theory of action*** explains how the intervention is constructed to activate the theory of change. By understanding the program theory, an evaluation can explore and assess the extent to which the causal processes are being activated and the intermediate steps of change are occurring.

Some programs have an additional element of theory, if the program is built around a specific social science theory about how change occurs. For example, a behavioural change program may be based on Bandura's social learning theory, or Ajzen's theory of planned behaviour.

Although programs are diverse across all areas of policy, many programs have a similar theory of change. Funnell and Rogers have used their work on program theory to describe five common '***program archetypes***'. These are 'carrots and sticks' (incentives and penalties), case management, community capacity building, direct service delivery, and information (providing advice, information or education). These archetypes are particularly useful for evaluators moving across sectors. For example, an evaluator with experience of programs using the information archetype in another sector will be able to bring that understanding to an energy efficiency program, because it uses the same theory of change.

Evaluation is sensitive to the context – of the program and of the evaluation

> *Context is a force in evaluation. It shapes our practice, influencing how we as evaluators approach and design our studies, how we carry them out, and how we report our findings. Context also moderates and mediates the outcomes of the programs and policies we evaluate* (Rog et al 2012).

Evaluation has become increasingly sensitive to context, because context shapes a program, its outcomes and the evaluation itself. In the case of energy efficiency, so much depends on the context. Energy use is social, contextual, and constrained by a range of factors both physical and invisible. (Lutzenhiser 2014, Shove 1998)

> *Programs work (have successful outcomes) only in so far as they introduce the appropriate ideas and opportunities (causal mechanisms) to groups in the appropriate social and cultural conditions (contexts).* (Pawson and Tilley 1997)

For example, a virus causes illness (a known causal mechanism), but the impact is not the same for everyone who is exposed. Some people have a stronger or weaker immune system, some people may be vaccinated, and some people become seriously ill. When we think about the non-energy benefits of energy efficiency programs, there is evidence that improving the insulation or heating in homes can improve people's health – but the context is key here. It has the potential to

[4] The definitions in this section come from the work of Professor Patricia Rogers and Sue Funnell (Funnell and Rogers 2011), which is being taken up internationally. There are other ways of describing program theory, and some people use the terms 'program theory' and 'theory of change' interchangeably. From my experience, the separation of theory of change and action is very useful for evaluation. If you want to understand program theory and its implications for program design and evaluation, this book is essential.

improve the health of people with chronic illness who live in poor quality homes in cold moist climates. In this case, cold indoor temperature is a mechanism that when added to the context of ill health and poor housing, can lead to worse health. A program that seeks to reduce this negative mechanism by adding insulation can - in the right context - lead to positive outcomes for people's health.

An evaluation is also shaped by the context – of the program, its implementation stage, the location, the political environment and the decisions to be informed by the evaluation (Julnes and Rog 2015).

The sensitivity to context in mainstream evaluation means that practitioners have learned to consider programs in terms of 'what works, for whom, it what situation', and to discern the level of outcomes that is reasonable to expect in a given context. This awareness of context – and its influence on program design, outcomes and evaluation – will be particularly useful for designing evaluation to consider both energy and non-energy impacts of energy programs.

Evaluation is interested in causality, and causality is sensitive to context

> *Causality is a central issue in making judgements about a program's merit, worth or significance. The classic causal question, in all its simple brilliance, is: Did the program produce the desired and intended results?* (Patton 2008)

Program theory explains the causal logic within a program, and causality is an important topic in mainstream evaluation. This is not as simple as looking for 'what works' and applying that program everywhere. Outcomes are dependent on context, and looking at causality helps us work out how the context of a program influences the causal processes and therefore the outcomes.

Nancy Cartwright (an English philosopher of science) differentiates between three types of causal claims that are relevant to evaluation and program planning (Cartwright and Hardie 2012).

- It works somewhere – based on findings from an impact evaluation
- It works in general – based on findings from synthesis of a range of impact evaluations
- It will work for us – is a question of judgement; the potential that it will work for us depends on the context in which it is implemented and the quality of implementation.

It is not uncommon for decision-makers to leap from 'it works somewhere' to 'it will work for us'. Then being disappointed when the program performs badly due to differences in context or implementation.

Different theories of change require different evaluation approaches to assess causality. Program designs and evaluations are increasingly noting the difference between necessary and sufficient conditions for causality (Cartwright and Hardie 2012, Funnell and Rogers 2011). This recognises that each aspect of a program may contribute to overall change, but that the greatest change comes through the necessary and sufficient conditions – either conditions of the program or conditions of the context in which it is implemented. Mainstream evaluation uses a range of methods to assess causality (such as contribution analysis), and this is currently an area of development in mainstream evaluation practice[5].

[5] Michael Scriven, a philosopher and major influence on evaluation, has written since the 1950s about causation from the perspective of philosophy of science, addressing Hume's problem of induction. He has returned to this topic in recent years in journal articles.

Conclusion and a call to action

Energy program evaluators must respond to the increasing expectations of energy policies and programs: to contribute to a range of policy outcomes, and for these programs to be evaluated to provide useful knowledge to decision-makers. The traditional technical focus of energy efficiency programs and evaluation is a distinct limitation when it comes to evaluating for the multiple benefits of energy efficiency. A new comprehensive approach to energy program evaluation will be required. While it is tempting to think that energy program evaluation can remain discrete and technical, I hope my description of the richness of mainstream evaluation gives you a sense of the potential gains from joining the mainstream.

To begin, you could join the professional society for evaluators in your country, read evaluation journals and some of the evaluation references in this paper. Attend local conferences and seminars on mainstream evaluation and become familiar with the diversity of evaluation theory and practice. There are many opportunities to learn. By participating with mainstream evaluation you can learn from their past mistakes and avoid 'reinventing the wheel'. Don't be put off by the different vocabulary, concepts and practices within mainstream evaluation. Explore the differences and seek to understand the differences behind the differences. Tell them about the challenges of evaluating for the multiple benefits of energy efficiency. Learn from each other.

Once you get connected with mainstream evaluation and understand what it has to offer, the next step will be to work together to develop a new and comprehensive approach to energy program evaluation, drawing on the expertise of both evaluation professions.

References

Alkin Marvin 2013. *Evaluation roots : a wider perspective of theorists' views and influences* (2nd ed), Sage Publications

Berriet-Solliec, Marielle, Pierre Labarthe and Catherine Laurent 2014. 'Goals of evaluation and types of evidence' *Evaluation,* vol. 20(2).

Blumstein, Carl, Seymour Goldstone and Loren Lutzenhiser 2000. 'A theory-based approach to market transformation', *Energy Policy,* vol. 28

Cartwright, Nancy 2011. 'A philosopher's view of the long road from RCTs to effectiveness', *The Lancet,* volume 377

Cartwright, Nancy and Jeremy Hardie 2012, *Evidence-based policy : a practical guide to doing it better.* Oxford University Press

Chelimsky, Eleanor 2015. 'Credibility, policy use and the evaluation synthesis', in Stewart Donaldson, Christina Christie & Melvin Mark (eds), *Credible and actionable evidence: the foundation for rigorous and influential evaluations* (2nd ed), Sage Publications

Davidson, E Jane 2005. *Evaluation methodology basics: the nuts and bolts of sound evaluation.* Sage Publications

Davidson, E Jane 2015. 'Question-driven methods or method-driven questions how we limit what we learn by limiting what we ask', *Journal of MultiDisciplinary Evaluation,* vol. 11(24)

Funnell, Sue and Patricia Rogers 2011. *Purposeful program theory: effective use of theories of change and logic models.* Jossey-Bass

Julnes, George 2012. 'Developing policies to support valuing in the public interest', *New Directions for Evaluation*, number 133

Julnes, George and Debra Rog 2015. 'Actionable evidence in context: contextual influences on adequacy and appropriateness of method choice in evaluation', in Stewart Donaldson, Christina Christie & Melvin Mark (eds), *Credible and actionable evidence: the foundation for rigorous and influential evaluations* (2nd ed), Sage Publications

Lutzenhiser, Loren 2014. 'Through the energy efficiency looking glass', *Energy Research & Social Science*, vol 1(2)

Mark, Melvin 2005. Evaluation Theory or What Are Evaluation Methods for? in *The Evaluation Exchange*, vol. 11(2). Retrieved from http://www.hfrp.org/evaluation/the-evaluation-exchange/issue-archive/evaluation-methodology/evaluation-theory-or-what-are-evaluation-methods-for

Moezzi, Mithra 2015. Numbers, stories, energy efficiency, in *Proceedings of the 2015 ECEEE Summer Study on Energy Efficiency*, ECEEE

Owen, John 2006. *Program evaluation: forms and approaches* (3rd ed). Allen and Unwin. [There is also a version for readers in Europe and the USA, referencing local context, which is published by Guildford Press]

Patton, Michael Quinn 2008. *Utilization-focused evaluation* (4th ed). Sage Publications

Pawson, Ray and Nick Tilley 1997. *Realistic evaluation*, Sage Publications.

Rog, Debra, Jody Fitzpatrick and Ross Conner (eds) 2012. 'Context: a framework for its influence on evaluation practice', *New Directions for Evaluation*, number 135

Schwandt, Thomas 2015a.'Credible evidence of effectiveness: necessary but not sufficient', in Stewart Donaldson, Christina Christie & Melvin Mark (eds), *Credible and actionable evidence: the foundation for rigorous and influential evaluations* (2nd ed), Sage Publications

Schwandt, Thomas 2015b. *Evaluation foundations revisited : cultivating a life of the mind for practice*, Stanford University Press

Scriven, Michael 2015. 'Demythologizing causation and evidence', in Stewart Donaldson, Christina Christie & Melvin Mark (eds), *Credible and actionable evidence: the foundation for rigorous and influential evaluations* (2nd ed), Sage Publications

Shadish, William, Thomas Cook and Laura Leviton 1991. *Foundations of program evaluation: theories of practice.* Sage Publications

Shove, Elizabeth 1998. 'Gaps, barriers and conceptual chasms: theories of technology transfer and energy in buildings', *Energy Policy*, vol. 26(15)

Yarbrough, Donald, Lyn Shulha, Rodney Hopson and Flora Caruthers 2011. *The Program Evaluation Standards: a guide for evaluators and evaluation users* (3rd ed). Sage Publications