

Building evaluation capacity: Experiences from the NSW Centre for Road Safety

Barnes^a, B., Wilkinson^a, R., Walker^a, E., Ma^a, A., and Fernandes^a, R.

^aNSW Centre for Road Safety, Transport for NSW

Abstract

The Centre for Road Safety (the Centre) is increasing its capacity to evaluate. The Centre has thus embarked on a process of placing evaluation at the core of how we do what we do. The paper outlines the evaluation practices we are seeking to foster at the Centre. The significant measures we have taken to foster these evaluation practices include:

- the creation of a Research and Evaluation Unit;
- the development of *Guidelines for evaluation practice at the Centre for Road Safety*;
- a custom-made, co-designed and co-facilitated two-day training course on evaluation;
- the development of templates and guidelines to support ways of working, which entail a commitment to prioritising and resourcing evaluation; and
- the management of a number of significant evaluations.

The paper also outlines the next steps on our evaluation practice agenda.

With these combined changes, in a short period of time, the Centre has moved to a more systematic practice of evaluation, where evaluation is core activity, expected for every new initiative, and considered from the outset of the design stage. This will further embed and consolidate an evidence-based approach to policy and program development.

Increasing evaluation capacity

At the Centre for Road Safety

The Centre for Road Safety is increasing its capacity to evaluate. The Centre recognises that evaluation plays a key role in improving Safe System initiatives and in meeting the safety targets set in the *NSW Road Safety Strategy 2012-2021*. Evaluation will help ensure that good policy intentions result in appropriate service delivery and that road safety is mainstreamed across the wide range of partner organisations. Evaluation will help assess the extent to which road safety expenditure is well-spent to achieve road safety outcomes. The Centre has thus embarked on a process of placing evaluation at the core of how we do what we do.

Across the NSW public sector

The Centre's decision to increase our evaluation capacity precedes, but also dovetails with, current efforts across the NSW public sector to promote evaluation. These efforts were motivated by the NSW Commission of Audit's *Final report: Government expenditure (2012)*, which found that the public sector does not routinely and effectively evaluate expenditure outcomes to investigate value for money.

In response, the *NSW Government Evaluation Framework* (2013) sets out how evaluation will become a core component of the way the NSW public sector works. Its five key points are as follows:

- Agencies are expected to evaluate their initiatives, both new and existing.
- The public sector needs to build capacity and capability to evaluate its initiatives.
- Evaluation plans are required by Cabinet's Expenditure Review Committee (ERC) for all new proposals.
- Agencies are required to submit and report on rolling 12-month schedules of initiatives to be evaluated over the forward financial year.
- Agencies are mandated to proactively and publicly release the findings of evaluations, unless there is an overriding public interest against disclosure.

These requirements are a transition to a future in which all government expenditure is allocated to an initiative, all initiatives are evaluated, and an initiative is re-funded past its review or termination date only if the evaluation demonstrates its ongoing relevance, effectiveness and value for money.

This is a major challenge for the public sector as a whole, and a significant change even for the Centre, which was identified by NSW Treasury as one of a very small number of agencies that currently have a high level of expertise in evaluation.

Across the country

While the *NSW Government Evaluation Framework* applies only to our jurisdiction, it is interesting to note how it ties in with the current zeitgeist, with major government reviews in other jurisdictions coming to similar conclusions about the need for increased efforts on evaluation.

The National Commission of Audit's report, *Towards responsible government: Phase two* (2014), is a more controversial document than its NSW predecessor in terms of its recommendations on government expenditure. Nevertheless, it comes to a very similar conclusion, that there is a pressing need for the Australian Public Service to improve practices for evaluating the effectiveness of government initiatives and agencies, with the solution including: annual schedules; the presumption of publication; and a systematic approach to using the results of evaluation activity, namely conclusions about the effectiveness of initiatives, as part of the standard funding process.

On the specific subject of road safety, the Road Safety Committee, Parliament of Victoria argued extensively for the importance of best practice evaluation in its *Inquiry into serious injury*, to inform the allocation of government resources across evidence-based policies and interventions.

The nature and types of evaluation

At the Centre for Road Safety, evaluation is defined as "a systematic and objective process to make judgements about *the merit or worth of one or more initiatives*, by comparing information on selected criteria *according to a set of explicit or implicit standards*".

The text in italics is critically important. It is why the word "value" lies at the heart of the term "evaluation". One of the founders of evaluation, in one of his most iconoclastic statements wrote "Bad is bad and good is good and it is the job of evaluators to decide which is which" (Scriven 1986, p. 19).

While that statement may be somewhat overwrought, there is truth in it nevertheless. Evaluation needs to do more than provide information, it needs to make a judgement on what that information means. As Jane Davidson (2012) puts it, evaluation must answer not only the descriptive “what’s so” question, what the outcomes are, but also the evaluative “so what” question, how good, valuable, or worthwhile the outcomes are. Evaluation then needs to lead on to address the action-focused “now what” question, which, as discussed later, requires collective input from the evaluator and key internal stakeholders.

The types of evaluation that are typically undertaken within the Centre, and the preferred terminology, are as follows:

- Outcomes evaluation – the part of evaluation that focuses on the changes caused by an initiative. Changes may relate to behaviours, attitudes, events or circumstances. Economic evaluation is a subset of outcomes evaluation that determines how cost-effective the initiative has been or the ratio of costs to benefits.
- Process evaluation – the part of evaluation that focuses on the content, implementation and activities. It assesses the extent to which the initiative has been designed and implemented effectively.
- Monitoring – regular reporting of key metrics relating to process or outcomes.

The Centre is fortunate to have its ultimate outcomes clearly and transparently set out in the *NSW Road Safety Strategy 2012-2021* to focus our work: a 30% reduction in both fatalities and serious injuries over the decade. Furthermore, the Centre greatly benefits from the crash and hospital data collected by government agencies that are direct and high quality measures of those outcomes.

The practice of evaluation we are seeking to foster

The evaluation practices we are seeking to foster at the Centre for Road Safety include:

- building evaluation into design and management up front;
- using program logic to guide evaluation (and program design);
- focusing evaluation studies and reports on a small number of key evaluation questions;
- using an evaluation framework as the key evaluation planning document;
- using mixed methods to provide the data by which to answer these questions; and
- subjecting our own evaluations to the same critical review as the initiatives they evaluate.

Building evaluation into design and management up front

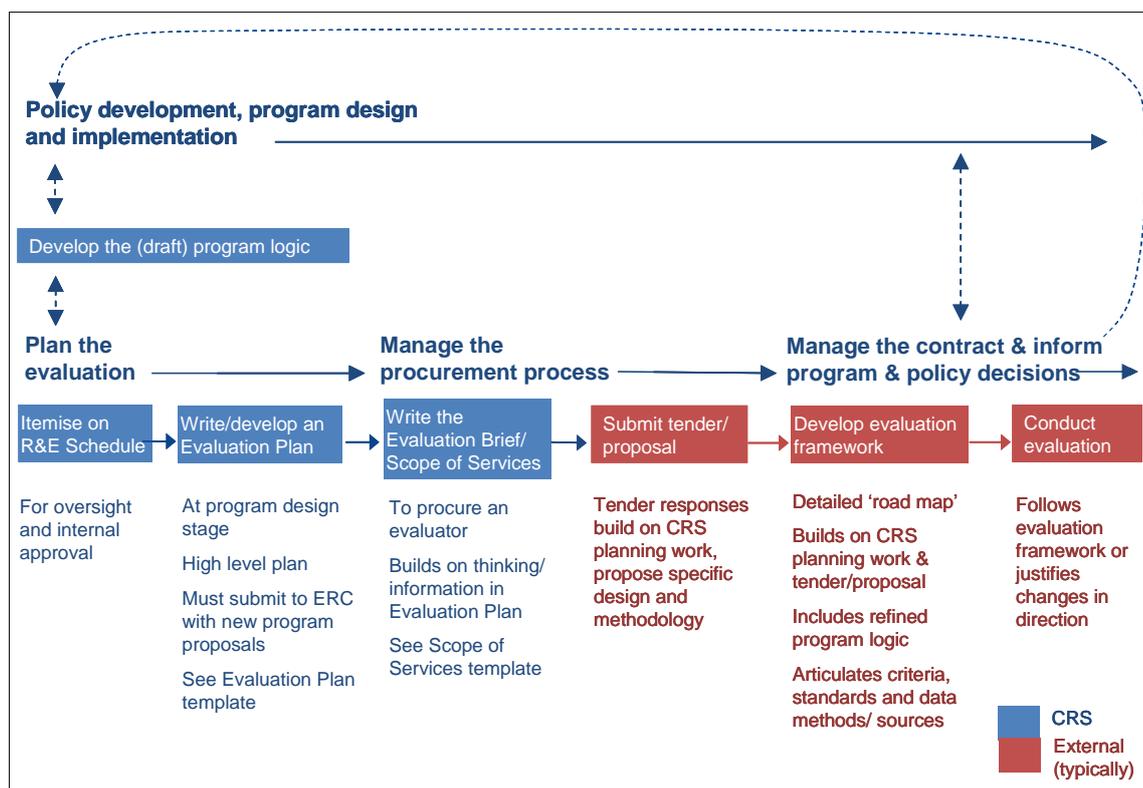
Evaluation planning commences at the design stage, to ensure evaluation thinking feeds in to the design of administrative data systems, as well as monitoring and reporting requirements, and to ensure that the requirements to support evaluation are set out in administrator and service provider agreements. This is not just about laying the groundwork for data collection for evaluation purposes; it is about improving initiative design and delivery as a result of thinking about evaluation. In this way, the progress and future of an initiative can be informed by:

- monitoring (the regular reporting of key metrics);

- process evaluation results, focused on the content, implementation and activities of an initiative; and
- outcomes evaluation results, as the data on behaviour change, crashes and the costs and benefits come to hand.

Figure 1 depicts the steps involved in planning and conducting an evaluation at the Centre, the role of program logic and a typical delineation of tasks between our staff and an external evaluator.

Figure 1. Evaluation planning and implementation



Using program logic to guide evaluation

A program logic is a flow chart or diagram that shows the process by which an initiative is supposed to contribute to its outcomes, and thereby to meet its objectives. It outlines the underlying need for the initiative, the key activities and resources, and then a chain of outcomes. It sets out what is predicted to occur, in a causal chain. It can be built up by drawing on available information, observation of the initiative in action, and/or stakeholder opinion.

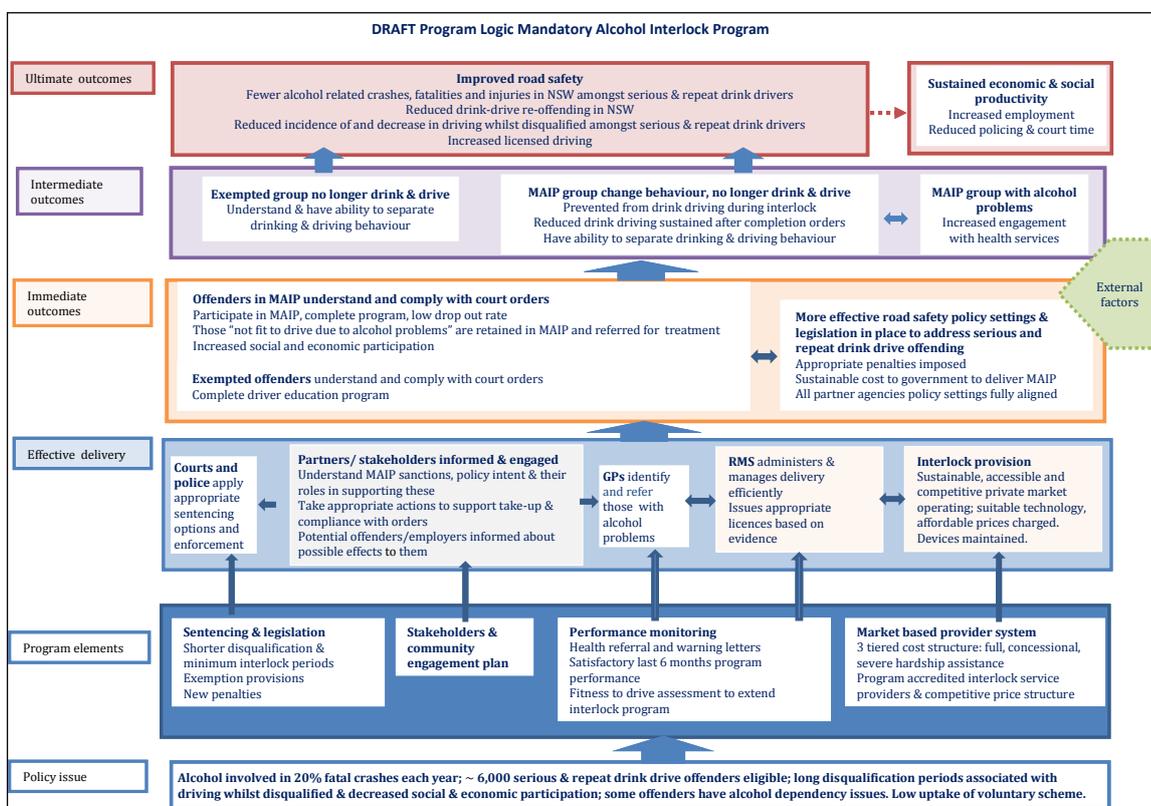
Program logic is not in itself evaluation, but rather a valuable design and management tool useful for evaluation. Developing a program logic as a first step will help focus the evaluation, and the initiative itself. The process of developing the program logic helps clarify the intended outcomes, facilitates a common understanding among stakeholders and highlights the strengths and potential weaknesses in the design of the initiative. It may even identify the need to revisit the program design, if it is evident that the activities cannot plausibly be linked in a causal chain through to the longer term outcomes. The diagram itself is also a useful communication tool. The program logic is often described as a “living” document which changes as an initiative is refined over time.

In terms of evaluation, program logic helps prioritise evaluation questions and the use of evaluation resources by highlighting the information gaps. It informs process evaluation by documenting the key elements of an initiative and their intended purpose. It informs outcomes evaluation by

articulating how the outcomes are expected to be achieved – these mechanisms then become the subject of investigation. If the initiative is not effective, the program logic will help distinguish implementation failure from theory failure. If the initiative is effective, the program theory will help determine what factors are instrumental to success and would need to be replicated in adapting the initiative to new situations (see Funnell & Rogers, 2011 for an excellent discussion of program logic).

The draft program logic for the Mandatory Alcohol Interlock Program (current at July 2014) is provided at Figure 2 below. External consultants ARTD led the development of this program logic in close consultation with the Research and Evaluation Unit and program team. It has been an iterative process, involving significant discussions about the nature of the Program and exactly what it aims to achieve. The evaluation framework has subsequently been built around this program logic.

Figure 2. Draft Program Logic for the Mandatory Alcohol Interlock Program



Focusing evaluation studies and reports on a small number of key evaluation questions

Evaluations need to be focused on a few, clear high level evaluation questions. The Centre typically draws on a list of question types that has been adapted from Jane Davidson (2012). These are used as a thought starter and then tailored to the specifics of the initiative in question, depending on the purpose of the evaluation.

- **Need (appropriateness):** Was the program – and is it still – needed? How well does it address the most important root causes? Is it still the right solution?
- **Process:** How well designed and implemented is the program?
- **Effectiveness/Outcome:** How valuable are the outcomes? To what extent has the program met its intended outcomes? What are the unintended or unexpected outcomes?

- Causal factors: What works best for whom, under what conditions, and why/how?
- Economic result: How worthwhile was it overall? Which parts of aspects of the program generated the most valuable outcomes for the time, money and effort we invested?
- Sustainability: How sustainable is the impact? How sustainable is the program itself?
- Lessons learned: What lessons have been learnt (or need to be learned) from this program?

Given the renewed emphasis on the particular economic question of value for money, our evaluations will need to tackle this question more effectively. In the past this has been done better for evaluations of engineering works, but less well for evaluations of behavioural programs.

It is worth stressing that the answers need to be evaluative – how good is the initiative, and how can it best be improved? Summarising the data is not enough.

An evaluation framework as the key planning document

At the Centre, the term “evaluation framework” refers to a detailed road map for the evaluation. The content of the framework may differ according to the nature of the evaluation, but typically it will set out the evaluation objectives and key questions, the criteria and standards to be used in answering the evaluation questions, and the data collection methods and other data sources that will generate relevant information. It will also provide an overview of the initiative and policy context, the relevant governance structures, ethical considerations, risks and mitigation strategies. The evaluation framework might be the first deliverable of a commissioned evaluation. Alternatively it may be commissioned as a separate piece of work, for example if the evaluation is expected to be complex.

Using mixed methods to provide the data by which to answer evaluation questions

We are very fortunate at the Centre to have access to high quality and detailed data on licensing, driving offences and crash outcomes. This gives us ample scope to use sophisticated quantitative methods to answer evaluation questions. However, we strongly promote the use of mixed methods, as the quantitative and qualitative data complement each other. The quantitative data is better for discovering what happened, and establishing and describing relationships among variables. The qualitative data, from road users and from stakeholder interviews and consultation, is especially useful for discovering why and how things worked the way they did.

Subjecting our own evaluations to the same critical review as the initiatives they evaluate.

Finally, the Centre is keen to promote informal evaluations of all evaluations conducted. In addition to assessing the quality of the evaluation product, this involves assessing the process. Feedback is sought from internal staff, the external evaluator and delivery partners as appropriate on the commissioning, management and implementation of the evaluation, and the extent the findings were used. The aim is to improve internal processes and to provide constructive feedback to all involved.

Changes that place evaluation at the core of how we do what we do

In order to place evaluation at the core of how we do what we do, the Centre of Road Safety has made a number of significant changes.

The creation of a Research and Evaluation Unit

In 2012, a functional review of the Centre led to the creation of a small Research and Evaluation Unit (three staff). Among other things, this Unit was tasked with developing and implementing evaluation guidelines to ensure that evaluation was used to assess and improve the road safety impact of road safety initiatives undertaken across the Centre. The first two authors of this paper, Ben and Rebecca work in this Unit, and both have extensive previous experience as evaluation consultants.

Most individual evaluations at the Centre are being managed by the team which owns the initiative to ensure evaluation is integrated with design and management, and informed by content experts. The role of the Research and Evaluation Unit is to provide specialist expertise and input across all stages of a range of evaluation projects, but particularly the planning and design stage. The Unit will manage only the largest and most complex evaluation projects. This will ensure that the expertise of the Research and Evaluation Unit is spread appropriately across all the Centre's evaluation projects.

A key function of the Unit is to develop evaluation capacity across the Centre by providing evaluation training and developing resources, templates and ways of working, as described in more detail below.

The development of Guidelines for evaluation practice at the Centre for Road Safety

An organisational document, *Guidelines for evaluation practice at the Centre for Road Safety*, has been developed to guide evaluation practice across the Centre. It articulates how evaluation will be undertaken and used within the Centre; identifies governance arrangements, roles and responsibilities; and sets out good practice principles, drawing and building on the *NSW Government Evaluation Framework*. The *Guidelines for evaluation practice at the Centre for Road Safety* also include a two year plan of action to develop evaluation capacity and embed new ways of working.

Evaluation training across the Centre

All staff require a good foundational understanding of evaluation to build evaluation thinking into their work. They require the skills to design, plan, commission and manage external evaluations, or to support others in this role. This is consistent with the expectation that all NSW public servants will have and utilise at least baseline evaluation skills, as set out in the *NSW Government Evaluation Framework*.

To this end, a custom-made two-day evaluation training course was co-designed and co-delivered by:

- Duncan Rintoul, an experienced evaluator and evaluation trainer, from his own firm Rooftop Social Research and Evaluation;
- Dr George Argyrous, from the ANZ School of Government (ANZSOG), who runs Executive Workshops on Evidence-Based Decision-making and Evaluation for Public Sector Managers, both of which played a role in setting the Centre on its current path for evaluation;
- Ben and Rebecca from the Centre's Research and Evaluation Unit.

The collaborative approach enabled existing "tried and tested" training materials to be used as a foundation, but then tailored to our needs and context. Complementary Centre-specific material was

developed, including weaving two current case studies throughout the training, and explaining our specific processes for planning, commissioning and managing evaluation. This approach also ensured alignment with recommended evaluation design options in guidelines specific to road safety, namely *An introductory guide for evaluating effectiveness of road safety treatments* (Austroads, 2012).

Over two full days, the training covered:

- an overview of evaluation, including the various types of evaluation and the need to focus evaluation on a few, clear questions;
- an introduction to program logic, which sets out in diagrammatic form how an initiative is supposed to contribute to its intended outcomes, and which is useful for both evaluation as well as design and management of the initiative in question;
- the practical steps involved in planning, commissioning and managing evaluations; and
- how to put it all into practice.

Participants gained firsthand experience drafting a program logic, writing key evaluation questions and identifying criteria, standards and data sources.

An off-site location was chosen to minimise non-essential work interruptions. The four presenters delivered the training to up to 16 participants. The high facilitator-participant ratio meant each small group had the dedicated support of a facilitator throughout both days. Intentionally, staff members from different teams were selected to work together, to bring different perspectives to the table and to promote collaboration. The delivery mode comprised a mix of presentations, facilitated discussions and small group work on a case study, which built across the two days. Brief homework tasks were set to engage people in the subject matter of the day. Feedback forms were collected from all participants on both days so that changes to the structure and content could be implemented immediately in response to feedback.

The development of templates and resources to support ways of working

Templates and resources have been developed to assist staff in commissioning and managing evaluations. Importantly, the Centre's template for a Policy Project Plan includes a section on monitoring and evaluation. This reinforces the expectation that all significant initiatives will be evaluated, and ensures that the project manager provides key details about how a project will be evaluated, and the process by which a monitoring, evaluation and reporting plan will be developed. This is also important in creating the expectation that resources (particularly staff time, but also funding) will need to be committed to evaluation.

There is a template for a high level Evaluation Plan that is now required for any new proposal submitted to the Cabinet's Expenditure Review Committee, which vets all proposals to Cabinet which have budgetary implications. The Evaluation Plan may only be two to three pages, but is enough to demonstrate that the evaluation has been considered appropriately and that there is a plausible path to a useful evaluation.

There is also a template for a Scope of Services, required to commission an external evaluation. The topics covered are broadly the same as in the Evaluation Plan, but by this later stage the level of detail that is provided is substantially greater, in order that tenderers obtain a good understanding of the parameters of the initiative and its evaluation. Importantly the focus in this document is to clearly articulate the evaluation purpose and questions, while leaving substantial room for the external evaluator to make the case for a particular methodology and data collection activities to

address the purpose and answer the questions. The Centre also recognises that in many instances, the evaluation methodology can only be determined through the development of an evaluation framework as the first deliverable. It is therefore necessary to build in flexibility in both the Scope of Services and the resulting contract to accommodate methodological changes. In some circumstances, we have found it works best to commission the evaluation framework as a separate piece of work.

There are also guidelines for structuring evaluation frameworks and evaluation reports, for managing consultants and for informally evaluating an evaluation.

Good practice examples of key evaluation documents are being collated and stored centrally along with a set of recommended externally developed resources.

The management of a number of significant evaluations

Several significant evaluations are underway, enabling new evaluation practices and processes to be put to the test.

An evaluation framework is being developed for the Mandatory Alcohol Interlock Program (MAIP) that is due to commence in February 2015. The drafting of the evaluation framework is happening in parallel with various elements of the design and development of the initiative so that each can inform the other. External evaluation consultants ARTD have been engaged to do work in close collaboration with the Research and Evaluation Unit and the MAIP team. The Research and Evaluation Unit will take over the further development of the framework and refine it over time as MAIP development work progresses.

The Centre has very recently engaged The Monash University Accident Research Centre, in partnership with Transport and Road Safety (TARS) Research, to evaluate the Safer Driver Course. The initial two year contract will focus on delivering a process evaluation, commencing monitoring activities and laying the groundwork for a subsequent outcomes evaluation. The process evaluation will be used to inform the continuous improvement of the Safer Driver Course.

Evaluations are also well underway for other initiatives, conducted by a mixture of road safety research institutions and private sector evaluation consultancies. This includes the Restricted P Pilot Program (Urbis), **bstreetsmart** 2014 (Inca Consulting), the Pilot of Pedestrian Countdown Timers (ARRB), the evaluation of a Smart Rest Areas Trial (TARS). An evaluation framework has also been developed for a pilot of the Risk Classification System (TARS). There is a schedule of evaluations expected to commence in the next year or so, including an evaluation of the Safer Roads Program.

Next steps

There are several things on the Centre for Road Safety's evaluation practice agenda for the next couple of years.

Procedures to ensure professional and ethical conduct

High on the list is the establishment of more detailed guidelines and standard operating procedures to ensure the professional and ethical conduct of evaluation. Research, evaluation and quality assurance exist on a continuum of activity. It is not always clear what level of oversight is necessary for evaluations done by the Centre, and in particular when review by a Human Research Ethics Committee is required. Academic research centres have their own procedures, and tend to have all these different activities reviewed, perhaps for better and worse. However, if a university partner is not involved, government agencies need to articulate what they perceive to be the appropriate level

of oversight. Some government agencies have their own ethics committee and this can introduce both benefits and limitations. The Centre is not covered by a particular Human Research Ethics Committee. We are aware that ethics committees exist that operate on a not-for-profit, cost-recovery basis that take on work from organisations not already covered by their own Committee and when the need arises, we will investigate the feasibility of that option.

A closely related piece of work is the clear articulation of privacy obligations as they apply in the specific context of the kinds of evaluation undertaken by the Centre.

Further training

Additional internal evaluation training is also planned for the coming year, with topics likely to be drawn from how to prepare a better Scope of Services, how to evaluate Evaluation Tenders; how to effectively review an evaluation framework and report; and how to undertake an internal evaluation. We will also investigate how best to introduce more time-poor staff to the basics of evaluation in a training course that lasts less than half a day, rather than the full two days used for the training rollout.

New learning

Over the next couple of years, our understanding will be enhanced of how best to conduct evaluations in an engineering context. Evaluation as a discipline has grown up in the human services context, such as health, education and community services. We are confident that much of the current understanding of evaluation will apply and that, for example, program logic will be as relevant to the evaluation of engineering initiatives. Nevertheless, experience is sure to yield further insights into such evaluations. Changes or additions may be required in how we understand evaluation to best incorporate the evaluation of engineering initiatives.

Evaluation reporting

Over the next two years, several evaluations will conclude or reach significant reporting milestones. It will be an opportunity to improve the wide sharing of evaluation results, in a manner tailored to the particular audience. It will also be an opportunity to ensure the evaluation reporting occurs in such a way as to help inform decisions about the future of initiatives and how they can be improved.

Of course, evaluation can often fall into the category of important but not urgent tasks for government agencies. It will be a shared responsibility to ensure that delays are kept to a minimum and do not prevent evaluation results from being delivered in a timely way to be used when decisions about future funding and/or improvement need to be made, and not just to contribute to the wider body of knowledge.

Another way of improving the usefulness and wider accessibility of evaluation reporting is to promote the preparation of reports in a digestible and succinct format. Report should be short, sharp and to the point and should be structured by the evaluation questions, or perhaps the structure of the initiative, but certainly not by the methods of data collection. We have suggested the following possible structure, but it will need to be put to the test in our context:

1. Executive summary (1-3 pages);
2. Evaluation objectives & key questions (2 pages);
3. Evaluation methodology summary (1-2 pages);
4. Program description (3-4 pages);
5. Answers to key evaluation questions (10 pages, drawing on all the evidence, explaining the evaluation criteria and linking directly to the evaluation framework);
6. Recommendations/conclusions (2-3 pages); and
7. Appendices (as long as required, including methodology, complete findings, instruments etc).

Another key issue here is how to translate evaluation findings into policy decisions and practice. Evaluation is focused primarily on the evaluative “so what” question (how good, valuable, or worthwhile the outcomes are), but it must also lead on to address the action-focused “now what” question. Formulating an answer to this question, including specific recommendations, requires collective input from key internal stakeholders in addition to the evaluators. It is generally unrealistic to expect evaluators to be sufficiently across the organisational history, politics and budgetary considerations to do this alone. Even more important is the fact that recommendations for action are so much more likely to be implemented if key decision-makers have themselves been involved in choosing that course of action. The evaluators may present the evaluation findings to commence that planning workshop. They may also facilitate the action planning components of that workshop. They may even document the planning decisions made in the final evaluation report. But the commitment to what needs to be done must be shared.

Evaluating the NSW Road Safety Strategy 2012-2021

Significant planning work will commence soon in preparation for evaluating the *NSW Road Safety Strategy 2012-2021*. A long range view is needed to ensure we have the necessary individual pieces of research and evaluation to inform a higher-level evaluation of the Strategy. As individual evaluations conclude or reach significant milestones, this presents the opportunity to identify common themes, strengths and weaknesses, and to synthesise findings for particular subgroups and categories.

Conclusion

With these combined changes, in a short period of time, the Centre for Road Safety has moved to a more systematic practice of evaluation, where evaluation is core activity, expected for every new initiative, and considered from the outset of the design stage.

As the NSW Commission of Audit stated:

The value of program evaluation is not just in the evidence base built. Setting systematic evaluations and forcing an environment of greater transparency is likely to influence performance and behaviours on the ground.
(NSW Commission of Audit, 2012, p.342)

This is particularly important in organisational contexts like our own, where those making policy are functionally separate to those charged with service delivery. In this increasingly common scenario, evaluation becomes an effective and civilised way of ensuring that good policy intentions result in appropriate service delivery and that road safety is mainstreamed across the wide range of partner organisations.

Of course, there is still much to do. Certainly, we value the independence and expertise to be found in both road safety research centres and evaluation consultancies. We feel that, working together, we will achieve this vision of placing evaluation at the core of how we do what we do.

References

Davidson, J. (2012). *Actionable evaluation basics: Getting succinct answers to the most important questions*. Smashwords: Los Gatos.

Funnell, S. C. and Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. John Wiley: San Francisco.

National Commission of Audit. (2014). *Towards responsible government: Phase two*. Commonwealth of Australia: Canberra.

NSW Commission of Audit. (2012). *Final report: Government expenditure*. NSW Government: Sydney.

NSW Government (2013). *NSW Government Evaluation Framework*. NSW Government: Sydney.

Road Safety Committee, Parliament of Victoria. (2014). *Inquiry into serious injury*. Government of Victoria: Melbourne.