

Rapid performance evaluation of government environmental science

Summerell, G., Booth, C.A., Leys, J., Wilson, K., Goggin, C.L., Riley, M., Young, J. and Elmes, D.

*NSW Office of Environment and Heritage, PO Box 5336 Wagga Wagga NSW 2650
Email: gregory.summerell@environment.nsw.gov.au*

Abstract: The Science Division in the NSW Office of Environment and Heritage (OEH) assessed the performance of 54 sub-programs comprising 144 projects run by the Division in 2015-16. The evaluation was undertaken within approximately six months and included assessment of science projects on climate change, pollution, land and water, biodiversity and community. Termed the Performance Evaluation Framework (PEF), the process used a rapid and rigorous methodology to collect consistent data on the sub-programs and projects to assess their appropriateness, effectiveness, efficiency and impact, against a set of benchmarks.

The Senior Leadership Team (SLT) of Science Division initiated the evaluation and were closely involved throughout the process. A Working Group consisting of two members of the SLT and representatives from across Science Division developed the PEF with support from a small project team and technical experts. The PEF Working Group principally used the then NSW Government Evaluation Framework (2013) to undertake a formative evaluation, informed by published literature.

The SLT and PEF Working Group developed 11 core criteria and nine supplementary criteria to assess performance. Each criterion had five levels to evaluate the: **Worth** (appraisal of appropriateness); and **Value** (appraisal of impact, effectiveness and efficiency). Overall evaluation was described as **Merit** and was determined by the combined consideration of **Worth** and **Value**.

For each of the 11 core criteria, five categories of benchmarks reflecting the different science deliveries we undertake were developed by the Working Group and reviewed by the SLT. The benchmarks categories were: 1. Excellence in delivery and use of relevant, rigorous, accessible, timely science (Project Type 1: **Enhanced Science Delivery**); 2. Foresighting and policy development (Project Type 2: **Inform Policy**); 3. Environmental program delivery (Project Type 3: **Program Delivery**); 4. Public environmental data, information, knowledge (Project Type 4: **Public Information**); and 5. The community acts on and contributes to environmental science (Project Type 5: **Community Action**).

After piloting the PEF on a subset of projects, it was implemented across all 144 projects in Science Division. Project Leaders entered data for each sub-program and project listed in the 2015-16 Science Division Business Plan, under supervision from independent working group members to ensure consistency in scoring and interpretation of questions. Quality assurance was provided by mandating references to evidence used to assign a score and the Project Leaders ranked the Merit of each project or sub-program as **acceptable to outstanding, intermediate** or of **less merit**. The ranking was performed by comparing the sub-program and project score against role-based benchmarks. Results were plotted to facilitate comparison and communication.

The PEF delivered comprehensive, rigorous and consistent information to the Science Division SLT so they could compare performance across 54 sub-programs (made up from the 144 projects). The evaluation guided future strategic directions as well as financial and resource sustainability. Performance against relevant benchmarks also gave clarity to Science Division staff about expectations for project performance. The PEF enabled the SLT to better articulate the value of Science Division work. The ability to provide clear information to the OEH Executive about the merit and performance of Science Division. In addition, to the comparative PEF scores, information was also provided on linkages to OEH goals, science achievements to date and a statement as to why the people of NSW should invest in each sub-program and project.

On reflection, the success of the PEF occurred because the SLT were integrally involved in this initiative, and the PEF Working Group. Project team and technical experts enabled it to be implemented in a practical way. The PEF process demonstrated that rigorous evidence could be efficiently collected and analysed, and effectively communicated in a timely manner to standardise dialogue and data to support decision making.

Keywords: *Environmental science impact, performance assessment, program evaluation, science sustainability, government science*

1. INTRODUCTION

The NSW Government has committed to a series of reforms that will increase transparency of expenditure on programs and provide a better understanding of the outcomes delivered. The recommendations of the NSW Commission of Audit (Final Report Government Expenditure, May 2012) highlighted that **evaluation is a key tool**:

- to support **evidence-based policy and decision making** in government
- to help government **learn and adapt to changing environments**
- **for communicating and sharing valuable information.**

The Commission concluded that effective and sustainable expenditure to a large part hinges on greater transparency of government expenditure, including robust and systematic program evaluation for both existing and future programs. In response to the Commission of Audit, the NSW Government introduced a program evaluation for non-capital funded programs, and from August 2013 required agencies to conduct evaluations in line with good practice principles in the NSW Government Evaluation Framework, which was replaced in January 2016 by the NSW Government Program Evaluation Guidelines (NSW Audit Office, 2016).

The NSW Office of Environment and Heritage (OEH) focuses on ensuring that the people of NSW have a healthy environment and are supported to access, protect and enjoy their natural and cultural heritage. Science Division undertakes work that supports OEH's legislative requirements and NSW Government and corporate priorities. Science Divisions specific programs are informed by agency-wide strategic plans called the OEH Knowledge Strategy (<http://www.environment.nsw.gov.au/knowledgestrategy>). The principles of Ecologically Sustainable Development inform or underpin much of the organisation's existing and newly developed legislative responsibilities.

Science Division developed a Performance Evaluation Framework (PEF) to deliver comprehensive and consistent information that enabled comparative evaluation of performance across diverse sub-program areas. It also guided future strategic directions as well as financial and resource sustainability (<http://www.environment.nsw.gov.au/news/oeh-science-division>).

In the context of environmental program assessments similar research impacts have been undertaken by Hall et al. (2003). They developed a systems framework which recognised that research programs needed evaluation systems to provide opportunities for learning for progressive change. Their assessment criteria considered institutional context and was used to provide institutional learning. Hall et. al. (2003) concluded the critical information around institutional context for research and development were:

- How research priorities emerge
- The role of various actors in the transfer of knowledge
- The relationships between different actors and factors that influence this relationship
- How research performance is evaluated and rewarded.
- How R&D is held accountable to different interest groups
- How knowledge is built up
- How organisations reflect and learn.

Horton (1999) illustrated similar findings where institutional learning was an inherent and intuitive process. They showed that research management strategies can evolve and progress rapidly when programs have explicit systematic learning objectives and procedures. Chambers (1981) went further to establish principles for undertaking rapid assessments which were considered in developing the PEF. The principles we followed were;

- Proportionate accuracy – recognising the degree of accuracy required to capture intelligence on an issue.
- Timeliness is a critical factor in effective appraisal – rigorous information must be available to the decision makers when they need it.

The Science Division PEF assessed the work of Science Division against four categories: appropriateness, effectiveness, efficiency (NSW Government Evaluation Framework, 2013) and impact. These categories provided a common language to communicate the value of government environmental science and allow for continuous improvement through assessment.

The NSW Government Evaluation Framework defines **evaluation** as a systematic and objective process to make judgements about the **merit** or **worth** of one or more programs usually in relation to **appropriateness, effectiveness and efficiency**. Impact which relates to effectiveness, was distilled as an additional performance category to strengthen the performance evaluation and better communicate this critical requirement of government environmental science. The definitions of these performance categories were:

- **Appropriateness:** the degree to which the project addresses OEH statutory functions or requirements, and/or aligns with NSW Government and OEH priorities.
- **Impact:** the degree to which the project will make a difference to OEH, our customers, or the people of NSW.
- **Effectiveness:** the degree to which the project will be rigorously undertaken to produce the right outcomes, with the outputs delivered to the right person at the right time.
- **Efficiency:** the degree to which the relationship between inputs and outputs is timely, cost effective and to expected standards.

2. METHODS

2.1. Governance

Under the guidance of the Executive Director of Science Division, the Senior Leadership Team (SLT) initiated the evaluation and made the final decision about how to proceed.

Primary development of the evaluation process was by a dedicated small project team with the support of a cross Division Working Group which included members from each of seven Branches and two members of the SLT. Working Group members were selected because of their pragmatism, knowledge and experience. Development of the collection, analysis and reporting tools was done by the project team, with input from the Working Group and technical experts. A method to capture accurate information quickly was proposed with categories, criteria, levels of performance and benchmarks finalised in collaboration with the SLT.

The Working Group wrote a guide for Science Division staff to help them with the PEF process. The guidance document included definitions and criteria for evaluating performance. The Working Group also supported project managers to complete data entry.

Technical experts assisted in designing and building the infrastructure to capture, analyse and present information from the 54 Sub-programs made up of 144 projects. Although PEF data was multi-relational (one sub-program to many projects) and would have been better captured using a relational database, Microsoft Excel 2013 was used because it aligned with OEH financial systems and is a common platform familiar to all staff. A total of 198 linked Excel workbook files were used to capture PEF data. For technical specifications of the PEF, see Young and Leys (MODSIM 2017).

The proposed methods were reviewed and approved by the SLT to ensure they met their criteria for speed, rigour, transparency, and effective communication to the OEH Executive. The involvement of the two members of the SLT on the Working Group was very important in terms of delivery and continuing acceptance and support by the broader SLT. Following successful piloting of the proof of concept with a small subset of projects the PEF was built to assess all 144 science projects. The iterative involvement of the SLT in the development of the PEF – and committing their Branches to doing it – was also critical to its success.

Quality assurance was provided by mandating references to evidence, to substantiate the selection from among the standardised five levels of performance for each of the 11 core criteria. Evidence included published papers, databases etc. This critical step increased rigour of the process, and therefore, confidence in the assessment. It was an important element in furthering evaluation capacity of participating Science Division staff.

2.2 Implementation

The SLT approved **11 core criteria** for compulsory and **nine supplementary criteria** for optional inclusion in the PEF. The supplementary criteria provided an additional dimension to the evaluation, including more detail and evidence to support scores for the core criteria.

Performance of each project was evaluated for each of the 20 PEF criteria against four areas: appropriateness, effectiveness, efficiency and impact. Each criterion had five levels of performance which enabled the **Worth** (appraisal of appropriateness); and **Value** (appraisal of impact, effectiveness and efficiency) of each program

and project to be determined. The overall performance was described as **Merit** and was determined by combining scores for **Worth** and **Value**.

One challenge for the PEF was to enable a comparison of scores for equivalent projects and sub-programs. Because Science Division has a diverse range of projects from governance to citizen science, to understanding biodiversity process, the SLT and Working Group identified a range of *project types* with a different benchmark. The project types were based on the **Five Roles of Science** (Summerell *et al.* 2015) which had been used successfully by the Executive Director of Science Division to communicate the nature and extent of OEH environmental science delivered to customers. The five Roles (with *Project Types* in brackets) were:

1. Excellence in delivery and use of relevant, rigorous, accessible, timely science (Project Type 1: **Enhanced Science Delivery**)
2. Foresighting and policy development (Project Type 2: **Inform Policy**)
3. Environmental program delivery (Project Type 3: **Program Delivery**)
4. Public environmental data, information, knowledge (Project Type 4: **Public Information**)
5. The community acts on and contributes to environmental science (Project Type 5: **Community Action**).

Role-based benchmarks were developed for each of the five Project Types by identifying an expected score for the 11 core criteria. For example, a project which is designed to ‘Inform policy’ (Project Type 2) tends to require large internal investment because the projects explore initial concepts. A large investment from external partners is not required. Therefore, the benchmark score for ‘Leverage’ is low. In contrast, a project designed for program delivery (Project Type 3) should already be well established, and attract significant co-investment to maximise service delivery across government. Therefore, the benchmark for ‘Leverage’ would be expected to be high.

To create the role-based benchmarks, the five levels for the 11 PEF core criteria were assigned a performance class **for the role**. Performance classes were:

- **Acceptable to Outstanding Merit (Green)**: projects with criteria scores considered to be of high merit in relation to the role, i.e. high worth and/or value with respect to the Category concerned. This was ranked as acceptable to outstanding performance.
- **Intermediate Merit (Amber)**: projects with criteria scores considered to be of ‘Intermediate merit’ in relation to the role, i.e. intermediate worth and/or value with respect to the Category concerned. This was ranked as intermediate performance
- **Less merit (Red)**: projects with criteria scores considered to be of ‘Less merit’ in relation to the role and were of low worth and/or value with respect to Category concerned. This was considered to be underperforming.

An example of the structure of the benchmark for ‘Enhanced Science delivery’ for the 11 core criteria, using the five levels of performance is provided in Table 1.

Table 1. The benchmarks of performance for ‘Community Action’ project.

Community Action		Acceptable to Outstanding Merit	Intermediate Merit	Less Merit
Appropriateness	NSW Legislation	5	2	0
	NSW Government/Ministerial Priority	5	1	0
	OEH Knowledge Strategy	5	2	0
Impact	Impact OEH	5	0	1
	Realisation NSW Community	5	2	1
	Scope	5	2	0
Effectiveness	Data and Information Sharing	5	3	2
	Risk	5	3	1
	Customer relationship	5	3	2
	Leverage	5	0	2
Efficiency	Efficient delivery	5	3	1

The use of the role-based benchmarks allowed comparison of equivalent sub-programs and projects. The performance of the sub-programs against each of the five roles were visualised in radar diagrams.

The performance of the sub-program (a rollup of multiple projects) called *Landscape information, data and knowledge* and its three component projects were assessed across 11 core criteria against its role in Community Action (Figure 1). As was the case for most sub-programs in Science Division, this sub-program contributes across multiple roles. To avoid unnecessary complexity, for each sub-program or project, up to three roles were assigned.

The contribution of this sub-program to ‘Community Action’ was 30% (Figure 1). The contribution of the sub-program to each role changes throughout its life. For example, a sub-program may emphasise informing policy when it starts, and focus on implementation once the policy is developed. The PEF process used different combinations of benchmarks for different stages in the life of a sub-program to recognise the changing roles as institutional context and learning develops.

The performance of the sub-program was ‘Acceptable to outstanding merit’ for seven criteria: NSW Government/Minister; OEH Knowledge Strategy; Impact OEH; Realisation NSW Community; Scope; Customer relationship and Efficient Delivery. Its performance was of ‘Intermediate merit’ for three criteria: NSW Legislation; Data and information; and Risk. Its performance was of ‘Less merit’ for one criterion: Leverage.

By comparison, the performance of Project 1. *Accessible and collaborative landscape data and knowledge management* (red line in Figure 1) was assessed as ‘Acceptable to outstanding merit’ for seven criteria. Its performance was of ‘Intermediate merit’ for NSW Legislation; Risk and Efficient delivery. Its performance was of ‘Less merit’ for one criterion: Leverage.

Project 2 titled *Modelling landscape status and trends* (green line) and Project 3 titled *Landuse and management assessments and impacts* (blue line) were of ‘Acceptable merit’ in relation to ‘Efficient delivery’. However, Project 1 was of ‘Intermediate merit’ for this criterion.

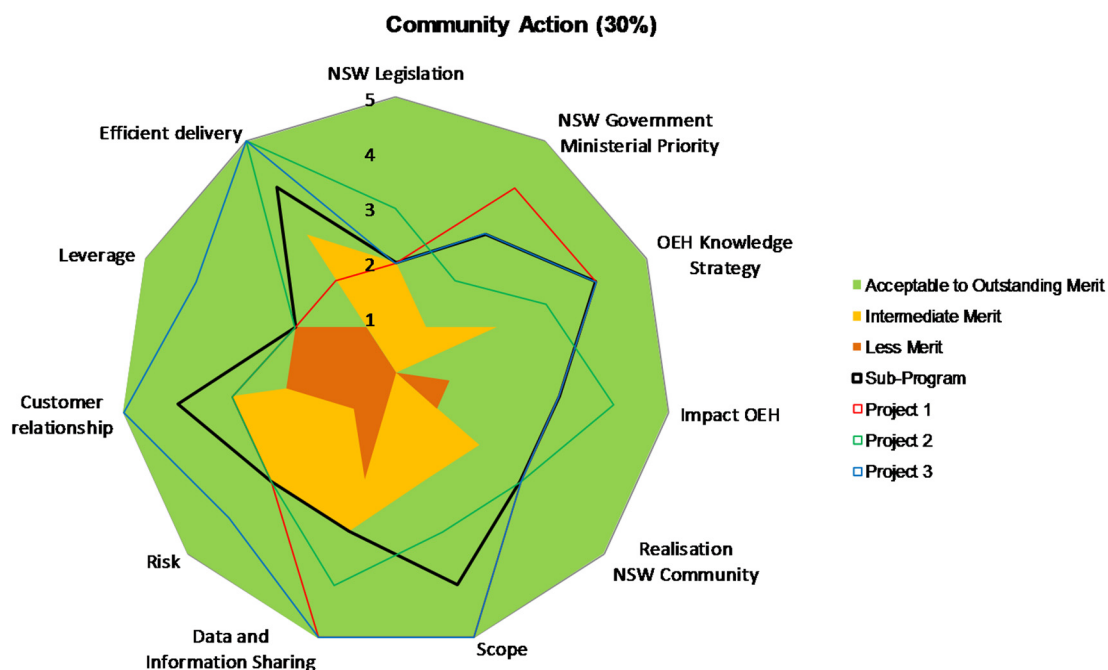


Figure 1. The radar diagram visualises the performance of the sub-program *Landscape information, data and knowledge* (black line) against 11 criteria for the ‘Community Action’ benchmark. The performance of the three projects comprising the sub-program are illustrated in red (Project 1), green (Project 2) and blue (Project 3). The benchmarks for ‘Acceptable to Outstanding merit’ are illustrated in green, ‘Intermediate Merit’ in amber, and ‘Less Merit’ in brown.

2.3 Evaluation

The SLT developed five questions to evaluate the performance of sub-programs. The questions were reported as text and graphs for the SLT (See Young and Leys, MODSIM 2017).

Q1 What is the spread, minimum, maximum, median and average investment of Science Division to the sub-program for the 2016-17? This assessed the expenditure by Science Division to the sub-program against expected investment.

Q2 What is the distribution of work (FTE, cash) by Science Division against the five roles? This describes the delivery of science against the five roles, showing where the Science Division attracts co-investment, informs policy, provides information for community decision-making, etc; and indicates skill sets and work effort.

Q3 What is the spend of the sub-program directed at each OEH goal (choosing only 1 goal for a sub-program to report to)? This question enables reporting on effort and the nature of services being contributed by Science Division to OEH goals.

Q4 Provide guidance on spectrum of performance. The PEF reported the top and bottom 10% of projects to identify areas for celebration or improvement. These results guided allocation of resources in 2016-17 (positively or negatively). This allowed mature conversations to develop such as are projects struggling with delivery because they are harder to deliver than the resources allocated (i.e. is there a problem of resource allocation not competence of delivery as people are delivering the best they can with the money allocated).

Q5: What investment does Science Division make in different partners? This information identifies partners in which Science Division is most heavily invested, as well as trends / anomalies for a strategic discussion.

The performance of all sub-programs was compared. This enabled each of the seven Directors in Science Division to use this information to discuss performance – and expectations - with Projects Leaders, in a consistent way across the Science Division.

3. CONCLUSIONS

The PEF collected rigorous evidence about the merit of the diverse sub-programs in OEH Science Division so their performance could be compared and their value articulated to decision makers. The PEF increased transparency of the subsequent prioritisation of future science effort. It also identified areas for improvement, especially around leverage of effort.

The PEF also clarified expectations of performance against benchmarks for Science Division staff, particularly about project delivery and leverage. To meet an immediate need of the SLT, the PEF additionally demonstrated and communicated the allocation of Science Division resources. The PEF results are now being used iteratively to improve performance of sub-programs and projects. The Division plans to continue using the PEF to continue improvement and consistency in sub-program evaluation. One component this method didn't address well (only 1 question) was "efficiency" of a program to conclude if a program may be performing well but is over resourced.

This PEF allowed the value of the work by Science Division to be better articulated and communicated to decision makers. In particular, it enabled Science Division to demonstrate its responsible use of resources to support NSW Government priorities. The PEF's comprehensive and comparative information informed dialogue between Science Division and the OEH Executive about the Division's delivery to OEH priorities. Indeed, a significant outcome of the PEF was that the value of the work of Science Division to the NSW Government could be visualised by the OEH Executive.

The PEF also enabled the value of government environmental science sub-programs and projects to be comparatively assessed and demonstrated potential to facilitate continuous improvement. These outcomes are important for government environmental science in the pursuit of Ecologically Sustainable Development. Government environmental scientists work at the interface of environmental issues, governance, business and communities. Further development and use of the PEF can help to optimise the benefits derived from their research, knowledge and often uniquely positioned long-term experience. Evaluating the performance of their sub-programs and projects in contributing to policy, planning and government decision making is pivotal to the delivery of government customer-focused environmental science for enhanced sustainability as outlined by (Summerell *et al.*, 2013; 2015).

On reflection, an important contributor to the success of the PEF was the active engagement of two SLT members in co-designing the process and championing the initiative, and the iterative involvement of the entire SLT. Executive commitment and leadership was also fundamental as was the support provided by the Working Group from across Science Division. Collaboration served to make the process targeted, applicable and practical for all staff. It also enabled the Project Team and technical experts to rapidly co-develop, pilot and implement the PEF to provide and appropriately draw on its comprehensive, rigorous and consistent information.

Another key facet in the PEF's success was that it enabled flexible benchmarking that recognised the different types of science projects. In this instance, this was represented by the five roles of science. The PEF presented rigorous consistent evidence in a timely manner that was visualised so it informed decisions of the SLT and OEH Executive. It also provided evidence of the merit of Science Division's effort to support the NSW Government priorities to sustain and protect the NSW environment.

ACKNOWLEDGEMENTS

This project would not have been possible without the support of the Senior Leadership Team of OEH Science Division, Klaus Koop, Jeremy Black, Tim Pritchard, Jo White, Alison Phin and members of the Working Group, Jo Ling, Angela McCormack and Richard Hicks. We thank all Project Managers who assessed their projects so a comparison of all projects was possible, and Amelia Swan for final proof reading of the manuscript.

REFERENCES

- Bella, D.A and Williamson, K.J (1986). Conflicts in interdisciplinary research. *Interdisciplinary Analysis and research*, 6, 347-354.
- Chambers (1981). Rapid Rural Appraisal: rationale and repertoire. *Public Administration and development* 1, 95-106.
- Hall A, Sulaiman V, Clark, N and Yoganand B (2003) From measuring impact to learning institutional lessons: An Innovation system perspective on improving that management of international agricultural research. *Agricultural Systems* **78**, 213-241
- Horton D (1999). Building capacity in planning, monitoring and evaluation: lessons from the field. *Knowledge, technology and Policy* 11 (4), 152-188.
- NSW Audit Office (2016) NSW Government Program Evaluation Guidelines, NSW Audit Office. http://arp.nsw.gov.au/sites/default/files/NSW%20Government%20Program%20Evaluation%20Guideline%20January%202016_1.pdf
- NSW Commission of Audit (2012) NSW Commission of Audit (Final Report Government Expenditure, May 2012) https://www.treasury.nsw.gov.au/sites/default/files/pdf/NSW_Commission_of_Audit_Final_Report.pdf
- NSW Government Evaluation Framework August 2013 (2013) Government of New South Wales, Sydney Australia. http://arp.nsw.gov.au/sites/default/files/NSW_Government_Evaluation_Framework_Aug_2013.pdf
- Summerell G., Leys, J and Wilson K. (2013). Customer-focused science and knowledge management for sustainability in New South Wales, Australia. 20th International Congress on Modelling and Simulation MODSIM, Adelaide, Australia, 1-6 December 2013. <http://www.mssanz.org.au/modsim2013/K3/summerell.pdf>
- Summerell, G., Leys, J., Booth, S., Oliver, I., Wilson, K., Littleboy, M. and Jenkins, B. (2015) Customer-focused science for environmental sustainability: a continuing case study from the NSW Government. 21st International Congress on Modelling and Simulation MODSIM, Gold Coast, Australia, 29 Nov to 4 Dec 2015 www.mssanz.org.au/modsim2015/
- Young J and Leys J (2017) A spreadsheet application for evaluating sub-programs and projects Merit, Worth and Value. 22nd International Congress on Modelling and Simulation MODSIM, Hobart, Australia, 3 Nov to 8 Dec 2017 www.mssanz.org.au/modsim2017/