Meta-evaluation

Case study – summative multiple meta-evaluation

Australasian Evaluation Society
International Conference 2009

Chris Milne

ARTD CONSULTANTS
strategy & evaluation

Overview

1. What is meta-evaluation
2. Case study – what and why
3. Our method
4. Findings
5. Lessons

What is meta-evaluation

I invented the term forty years ago … and I use the hyphen

Michael Scriven:

"The evaluation of evaluations"

Systematic reviews of evaluations to determine the quality of their processes and findings

"Peer review for evaluators"

ARTD

What is meta-evaluation

Two senses:

a. Quality of an evaluation(s)

b. Synthesis of findings from evaluations
   - systematic review
   - meta analysis

ARTD

What is meta-evaluation

Program Evaluation Standards

Accuracy A12:

The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses. (Joint Committee, 1994, p. 185)

ARTD

Forms of meta-evaluation

Formative – the cook

Summative – the guests

ARTD

Types of meta-evaluation

	Single	Multiple
Formative	*Evaluators do this*	
Summative	*Commissioners do this*	*We did this*

ARTD

Criteria for meta-evaluation

Daniel Stufflebeam:

Program evaluations metaevaluation checklist 1999

Formative or summative

Uses Program Evaluation Standards 1994

30 standards x 6 checkpoints each (180!)

ARTD

Program Evaluation Standards 1994

Utility - information needs of intended users.

Accuracy –technically adequate information about the features that determine worth or merit.

Feasibility - realistic, prudent, diplomatic, and frugal.

Propriety - conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

http://www.wmich.edu/evalctr/jc/

ARTD

Utility – serves information needs of intended users

U1 Stakeholder Identification
U2 Evaluator Credibility
U3 Information Scope and Selection
U4 Values Identification
U5 Report Clarity
U6 Report Timeliness and Dissemination
U7 Evaluation Impact

ARTD

Accuracy – technically adequate information about the features that determine worth or merit.

A1 Program Documentation
A2 Context Analysis
A3 Described Purposes and Procedures
A4 Defensible Information Sources
A5 Valid Information
A6 Reliable Information
A7 Systematic Information
A8 Analysis of Quantitative Information
A9 Analysis of Qualitative Information
A10 Justified Conclusions
A11 Impartial Reporting
A12 Metaevaluation

ARTD

Feasibility - ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

F1 Practical Procedures
F2 Political Viability
F3 Cost Effectiveness

ARTD

Propriety - conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

P1 Service Orientation
P2 Formal Agreements
P3 Rights of Human Subjects
P4 Human Interaction
P5 Complete and Fair Assessment
P6 Disclosure of Findings
P7 Conflict of Interest
P8 Fiscal Responsibility

ARTD

Meta-evaluation in practice

How much, what sort?

2009 scan of meta-evaluations:

- Single meta-evaluations, both formative & summative
- Identified as meta-evaluation or m/e audit

18, all US

Leslie J. Cooksy, Valerie J. Caracelli *Metaevaluation in Practice: Selection and Application of Criteria* Journal of MultiDisciplinary Evaluation Vol 6, No 11 (2009)

ARTD

Meta-evaluation in practice

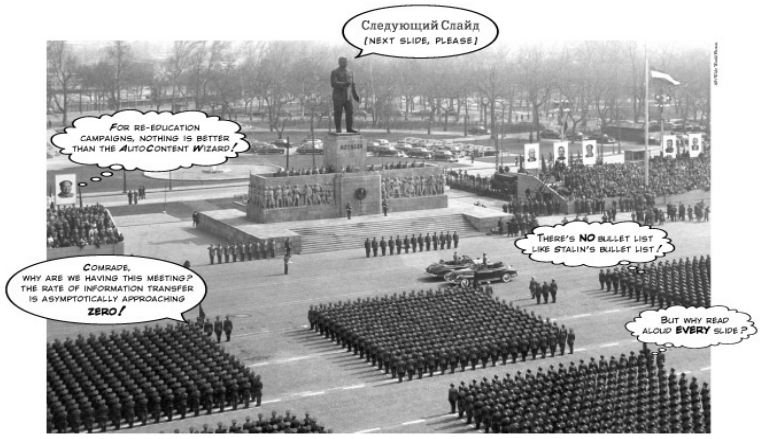
Metaevaluation Method by Metaevaluation Criteria ($N = 18$)

Method	Metaevaluation Criteria			
	Emergent	PgES	Tailored	Trustworthiness
Narrative reviews	7		1	
Checklists		4		
Semi-structured reviews		1	1	
Evaluation audits			1	3

Leslie J. Cooksy, Valerie J. Caracelli *Metaevaluation in Practice: Selection and Application of Criteria* Journal of MultiDisciplinary Evaluation Vol 6, No 11 (2009)

ARTD

Death by Powerpoint?



Edward Tufte, *The Cognitive Style of PowerPoint*

RTD

Case study - meta-evaluation of 11 evaluation reports

NSW Government Aboriginal specific programs

- Audit identified 150 - 2007
- “Evaluation reports” requested from agencies
- 60 documents

ARTD

Case study - meta-evaluation of 11 evaluation reports

Stakeholders – central agencies, DAA (steering group)

Objectives

- assess the quality of the evaluations
- examine the extent that they provided suitable evidence to support continuation of the individual programs esp appropriate, effective, efficient and value for money

Purpose

- improve quality of evaluation across sector

ARTD

Diverse programs and evaluations

Program	Agency and year of evaluation	Program status at time of evaluation	Size	Type of intervention	Focus of evaluation Internal or external evaluator
Circle Sentencing	Attorney General's 2003	New, year 1	Small One site, pilot	Apply model from elsewhere	Implementation Internal
the School in Partnership Initiative	DET 2007	New, year 1	Medium 10 schools	Emergent	Development, implementation Mixed
NSW Aboriginal Vascular Health Program 2000-2003	NSW Health 2004	Developing, year 3	Medium ($1-$3 m) 9+ sites	Emergent	Development, implementation External
Intensive Family Based Services	DOCS 2008	Established year 4	Medium ($1-$3 m) 2 sites	Apply model from elsewhere	Implementation, client outcomes, economics Internal
Aboriginal Client Services Specialist Program	Attorney General's 2005	Established 10 years	16 sites Medium ($1-$3 m)	No model	Program framework and future directions External

ARTD

Approach to this meta-evaluation

- Summative
- 11 evaluations
- Solely on evaluation reports ("product evaluation")
- Determine criteria and then apply it
- Iterative – steering group
- Individual assessment template
- Ratings to show overall pattern
- Summary report across all the evaluations

ARTD

Method

Design	Brief scan of other meta evaluations Confirm sample of "evaluation reports" Framework for criteria - Workshop to criteria Review and sign off
Apply criteria	Design and pilot template report Develop rating scale Prepare individual assessments (2-3 pages) Calibrate assessments Validate assessments - independent reviewer Individual assessments to steering group
Report	Confirm approach Analyse patterns and trends across individual assessments Prepare draft overall report, de-identified Provide agencies with draft overall report plus their individual assessment report Incorporate feedback into final report

ARTD

Two sets of Criteria

1. Quality of evaluation
2. Policy relevance – evidence to support continuation of the program

Reflect commissioning and conduct of evaluation

ARTD

1. Criteria for quality of evaluation

- Simplified Stufflebeam's checklist (15 standards, qualitative scale)
- Included specific attributes for programs for Aboriginal people eg

1. Utility	
Stakeholder identification	The range of stakeholders in the evaluation is identified in the information needs and potential contribution taken into account [specifically, identification of partnerships and Aboriginal government and community-based organisations in the relevant field)
2. Feasibility	
Appropriate stakeholder participation	Key stakeholder groups are represented in the evaluation process [specifically appropriate involvement of Aboriginal people eg reference group, balance of professional and local knowledge, capacity building for Aboriginal communities]
3. Propriety	
Respect for diversity	Perspectives of diverse groups are given due prominence, [specifically of relevant Aboriginal communities and people]
Disclosure of findings	Right to know audiences are identified and receive appropriate information on findings [specifically accountability to Aboriginal communities through appropriate feedback]

ARTD

2. Criteria for policy relevance

Whether the program:

- meets priority needs of Aboriginal people (appropriateness)
- delivers improved results for clients (effectiveness)
- has timely and cost effective service delivery processes (efficiency)
- is a reasonable use of resources (value for money)

Appropriateness – defined by current policies eg *Two Ways Together:*

the rights of Aboriginal people to determine the direction of their social, economic and political development

ARTD

Template report – summary section

High quality

Department of ABC – Evaluation of XXXXXX Pilot Program 2006	
The assessment	
The program	This program aims to xxxxxx
The evaluation	The Department commissioned an external evaluation to assess the effectiveness of the pilot in impacting on Aboriginal young people … It considered the effectiveness of the partnership and coordination between the parties in the pilots, and the appropriateness, effectiveness and efficiency of the processes.

Summary	
A. Quality of evaluation	This was a high quality evaluation across all four areas of utility, accuracy, feasibility and propriety, although there are some questions about the adequacy of resources for the review, and how it dealt with ethical issues. Its focus on processes and context was appropriate for a pilot program in a difficult context.
B. Relevance to current policy	The program addresses a high priority need for Aboriginal people in the two communities. The review provides initial evidence of largely effective processes, and indicates that the program partners can work together and that the program can have a positive impact with some young people. However it also identifies significant constraints particularly in the social and economic context of the two communities. It does not assess value for money. It provides a reasonable evidence base for decisions about refining and/ or further trialling the program.

ARTD

Template report – summary section

Mixed quality

Summary	
A. Quality of evaluation	The evaluation has a low level of utility with a poor report. It has a medium level of accuracy, with appropriate design and methods but shortcomings in presentation of data, and some conclusions that are not fully justified from the evidence. The evaluation had a high level of feasibility as part of the strategy itself. It had a medium level of propriety, with appropriate ethical approvals and inclusion of perspectives of local Aboriginal stakeholders. However the report had positive bias, and there was no indication of disclosure of findings to Aboriginal communities.
B. **Relevance to current policy**	The program is addressing a high priority need for Aboriginal people, although established prior to TWT. While the program engaged substantial numbers of Aboriginal women, evidence on effectiveness and efficiency is inconclusive, and the evaluation did not assess value for money. This limits the value of the evaluation for making decisions about continuation of the program. However, the outcome measures and data collection systems developed for the program should allow more conclusive assessments in future years.

ARTD

Template report – utility

A. Quality of the evaluation		
1. Utility		**High**
Addresses objectives	The objectives of the evaluation as described in the report were to • Xx • Yy The evaluation fully addresses these objectives.	High
Stakeholder identification	The evaluation identified and consulted with a wide range of stakeholders in the two communities.	High
Credible evaluators	The external review was by a consultancy firm (XYZ Consulting) with a record of review and evaluation in human service areas with the NSW and Australian governments, and includes an Aboriginal consultant.	High
Quality report	The report is clearly presented and effectively communicates the evaluation. It is structured around the evaluation objectives. It has a clear executive summary, and develops recommendations and future directions. While the methodology is described, there is little information on its limitations or confidence in the data.	High

ARTD

Template report – accuracy

2. Accuracy		**Medium**
Adequate program description	The report describes the background and rationale for the program, but does not clearly set out the program objectives, or the target level for funding or numbers of projects....	Medium
Appropriate design, methods, analysis	The methods, and data sources were appropriate (but may not have been feasible in the timeframe)... Some limitations of the data are identified... Analysis of the data is limited. There appears to be little systematic synthesis of data from the project reports or analysis of the pattern of successful applications ...	Medium
Effective data presentation	At a number of points the presentation is difficult to follow or relate to the evaluative arguments.... Quantitative data tables would be improved by the addition of percentages .... . While the report presents some summary quantitative data as tables, there is limited data reduction or systematic summaries of qualitative or quantitative data...	Medium
Justifiable conclusions	It is difficult to interpret a number of conclusions,, because the report does not identify whether the results are as expected, better than expected or worse than expected. Some conclusions and recommendations ... are justified from the data that is presented. However at a number of points there appears to be no clear link between claims and the evidence that is presented... Other conclusions do not appear to be justified. For example, ... the exec. summary refers to "outstanding projects that have realised exceptional value for money", but does not provided evidence to support this.	Low

ARTD

Template report – feasibility

3. Feasibility		Medium
Practical procedures	The methods appear to have been practical and implemented reasonably successfully. The evaluation drew on data from program monitoring, although in practice there were many gaps. Visits to projects in Aboriginal communities were properly included in the methodology, but attempting to make up to 40 visits within a short time frame was unlikely to be practical, and this is what the evaluators experienced. Visits to a sample of projects may have been more practical.	Medium
Appropriate stakeholder participation	The report does not identify whether there was a reference group for the evaluation. While the evaluation consulted a range of Aboriginal stakeholders, there do not appear to be processes of accountability such as presentation of findings to Aboriginal stakeholders to verify accuracy of data collected and conclusions, or opportunities for capacity building.	Not known
Reasonable resources	The budget and other resources for the evaluation are not identified. The evaluation was conducted over two months which, as the report acknowledges, was a major constraint, particularly in visiting numbers of Aboriginal communities.	Medium

ARTD

Template report - propriety

4. Propriety		Med
Balanced report	The report ... provides a relatively positive assessment, with some conclusions not fully substantiated (above). Unintended outcomes are not identified.	Low
Ethical approach	Ethics approval for the evaluation was obtained from [the peak Aboriginal body and sponsor government agency].	High
Respect for diversity	As indicated, the evaluation consulted a range of Aboriginal stakeholders at each site. However any distinct perspectives of Aboriginal policymakers or of different Aboriginal communities are not identified.	Medium
Disclosure of findings (where appropriate)	The evaluation was accountable to Aboriginal communities through the published report. The report does not refer to any arrangements to provide feedback in other forms to the local communities, although this may have occurred.	Medium

ARTD

Template report – policy relevance

B. Relevance to policy focus: to what extent does the evaluation show that the program contributes to the government's current policy focus		
Appropriateness		
Extent program addresses priority needs of Aboriginal people as identified in government policy, and specifically priorities under *Two Ways Together*.	The program comes under State plan priority F1 "improved health, education and social outcomes for Aboriginal people" and addresses the objective: safe families. It addresses one of the *TWT* priority areas for action: Families and Young People.... The program reflects *TWT* principles for improving agency capacity to work with Aboriginal people, and a degree of local planning and decision-making.	High
Effectiveness		
Extent program delivers improved results Extent program data identified clients, impacts and outcomes	The program engaged over 150 families .... The evaluation demonstrates positive client outcomes ..... The model and outcomes were described in sufficient detail to warrant decisions about replication. The program had suitable systems to collect client data on impact and outcomes. Some recommendations for improvements were made.	High
Efficiency		
Extent service delivery processes are timely and cost effective	The evaluation systematically assessed service delivery processes. It determined unit cost data ... but does not compare this with comparable programs. It makes suggestions to improve service levels and reduce costs.	High
Value for money		
Extent program is a reasonable use of resources	The cost benefit analysis concluded that the benefits outweigh the costs with a cost benefit ratio of 1.8. It would provide the basis of a business case for extending the program to suitable locations.	High

ARTD

Ratings

Table 4.2 Quality of evaluations overall and across the four quality areas

Program/ Agency	*Utility*	*Accuracy*	*Feasibility*	*Propriety*
High quality overall				(H = High, M = Medium, L = Low)
A	H	H	H	H
B	H	H	H	H
G	H	H	H	M
Medium quality overall				
H	H	M	H	M
J	L	M	H	M
K	L	M	M	M

ARTD

Findings – quality overall

- Two thirds high quality, esp utility and accuracy
  ⇒ Confidence in using for decisions
- One third medium quality
  ⇒ Limit value and credibility
  ⇒ Uncertainty (+,-)
- No pattern across type of program, agency or evaluation

Quality of the evaluations (number and % of evaluations)								
	High		Medium		Low		All	
Utility	8	73%	1	9%	2	18%	11	100%
Accuracy	7	64%	4	36%			11	100%
Feasibility	10	91%	1	9%			11	100%
Propriety	2	18%	9	82%			11	100%
Overall	7	64%	4	36%			11	100%

Findings – utility

Most high

- addressed needs of commissioning agency, but not always other stakeholders eg central agencies
- most evaluators credible whether "internal or external"

Two low – poor reports, unclear evaluation objectives

- two evaluators – competency in evaluation?

Actual utilisation not assessed

Quality of the evaluations (number and % of evaluations)								
	High		Medium		Low		All	
Utility	8	73%	1	9%	2	18%	11	100%
Accuracy	7	64%	4	36%			11	100%
Feasibility	10	91%	1	9%			11	100%
Propriety	2	18%	9	82%			11	100%
Overall	7	64%	4	36%			11	100%

Findings – accuracy

Two thirds high quality ->confidence in using findings

Others:

+ most had appropriate designs and methods, but
- limited analysis, poor data presentation
- conclusions not justified
- poor description inc expected outcomes or program logic

Quality of the evaluations (number and % of evaluations)

	High		Medium		Low		All	
Utility	8	73%	1	9%	2	18%	11	100%
Accuracy	7	64%	4	36%			11	100%
Feasibility	10	91%	1	9%			11	100%
Propriety	2	18%	9	82%			11	100%
Overall	7	64%	4	36%			11	100%

ARTD

Findings – propriety

Only two high, at least as documented in their report

Others did not report on:

- addressing ethics, or
- disclosing findings to key audiences
- esp to Aboriginal communities (as in *TWT*)

Quality of the evaluations (number and % of evaluations)

	High		Medium		Low		All	
Utility	8	73%	1	9%	2	18%	11	100%
Accuracy	7	64%	4	36%			11	100%
Feasibility	10	91%	1	9%			11	100%
Propriety	2	18%	9	82%			11	100%
Overall	7	64%	4	36%			11	100%

ARTD

Findings – policy relevance

The evaluation reports showed :

- all programs were appropriate (+)
- programs varied in effectiveness
- little information about value for money of the programs (?)
- few evaluations considered efficiency, cost, cost effectiveness or VFM (-)

The meta-evaluation pointed to
- areas for improvement, or
- negotiation around expectations eg more inclusion of cost-effectiveness

Policy relevance of the programs (number and % of programs)

	High		Promising *		Medium		Low		Not known		All	
Appropriate	11	100%									11	100%
Effective	1	9%	4	36%	2	18%	1	9%	3	27%	11	100%
Efficient	1	9%			3	27%	1	9%	6	55%	11	100%
Value for money	1	9%					1	9%	9	82%	11	100%

*Note: this rating was only used for effectiveness

ARTD

Quality of the meta- evaluation

Utility Information needs of intended users.	+ addressed objectives, identified stakeholders, credible evaluators, clear report → basis for conversations → suggested improvements → scoping sector-wide guidelines

ARTD

Quality of the meta- evaluation

Accuracy Technically adequate information about the features that determine worth or merit.	+ description, design, data presentation, conclusions + included agency feedback - reliance on "final" reports which were not always final or included all features - process use, utilisation not considered - limited confirmation of info by agencies ? validation ? rating and counting not standardised ? double jeopardy →caution with individual assessments → suited purpose (improvement)

ARTD

Quality of the meta- evaluation

Feasibility Realistic, prudent, diplomatic, and frugal.	+ practical, no burden on agencies + economic + agencies invited to respond - agency participation limited
Propriety Conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results	+ fair and balanced + evaluator interest identified + programs and individuals not identified ? no public disclosure of individual assessments

ARTD

Lessons for summative meta-evaluation

- more feasible just using reports as data, needs only reasonable resources
- but less accurate?
- could prompt improvements in reporting
- requires judgments drawing upon evaluation experience
- useful method for the purpose (improvement)
- process and results generate valuable conversations about evaluation

ARTD