What are the characteristics of *(good)* criteria?

Investigating the requirements for defensible public sector evaluation

Mathea Roorda
AEA, Denver
October 2014

About me…

One week out of every four I commute from Wellington, New Zealand's capital, to my home in the Marlborough Sounds, a remote area with about 15 permanent residents. Over the 13 years I have lived here I've become increasingly passionate about, and involved in, restoration of the flora and fauna in our Inlet.

I am as passionate about my paid work as I am about the place where I live. I am an independent evaluator with 15 years public sector evaluation experience. Working respectfully, and doing evaluation well, is important to me.

In 2010 I led a team that won the Australasian Evaluation Society's Best Evaluation Study Award for a two-year evaluation of a Pacific Islands-New Zealand labour initiative. I'm in the early stages of Ph.D. study at Melbourne University. My thesis topic is 'Developing defensible criteria for public sector evaluation'.

Mathea Roorda

mathea@evalueresearch.co.nz

www.evalueresearch.co.nz

Scriven's Logic of Evaluation

Establish criteria

Construct standards

Measure performance

Arrive at a judgment

In May 2014, as part of my Ph.D. study, I conducted an online survey of evaluators. All respondents had at least 5 years experience and were involved in managing or doing public sector evaluations. The majority worked either in a government agency or private practice. Respondents were asked to rate 17 sources for identifying relevant criteria for public sector evaluation. These sources were identified from a review of academic and grey literature.

137

evaluators in **Australian** and **New Zealand** public sectors

For each source, respondents were asked:

Which sources do you commonly **use**/consider **important**?

What influences your selection of sources?

What makes a source more important?

The next two slides look at the findings on respondents' use of, and their views about the importance of, the sources identified in the literature.

The 17 sources are listed down the left hand side of the chart. The percentage of respondents who say they 'often' or 'always' use a source is in green, and 'seldom' or 'never' in red. Likewise, in the chart on importance, 'important' or 'very important' is depicted by the green bars, and 'less important' or 'unimportant' are red. For ease of readability the data labels are excluded for percentages smaller than 10%.

The findings suggest respondents commonly use a range of sources when they are developing criteria of merit. Almost all respondents said they use program or policy objectives to develop their criteria and they consider this source to be important. The two sources least likely to be used (or considered important) are the political ministers and lobby groups. Needs assessment is a source where there is divergent use: 57 percent said they often/always use it; while 43 percent said they seldom or never use a needs assessment when developing criteria.

Policy objectives is the no. 1 source used to develop criteria (and it's also considered most important)

	Always	Often	Seldom	Never	USE
Program/policy objectives	86%	12%			
Policy documents	55%	41%			
Content expertise	37%	59%			
The government's strategies	41%	53%			
Program staff	55%	38%			
Literature on 'best practice'	38%	50%	12%		
Evaluation commissioner/sponsor	56%	30%	10%		
Stakeholders affected by program	53%	33%	12%		
Policy staff	42%	42%	16%		
Evaluation contract manager	45%	32%	15%		
Evaluator/s	39%	37%	20%		
Ethical guidelines or standards	46%	29%	23%		
Cultural expertise	25%	50%	23%		
Government legislation, regulations	21%	43%	34%		
Needs assessment	12%	45%	39%		
Political Minister	18%		50%	28%	
Lobby groups	18%		55%	25%	

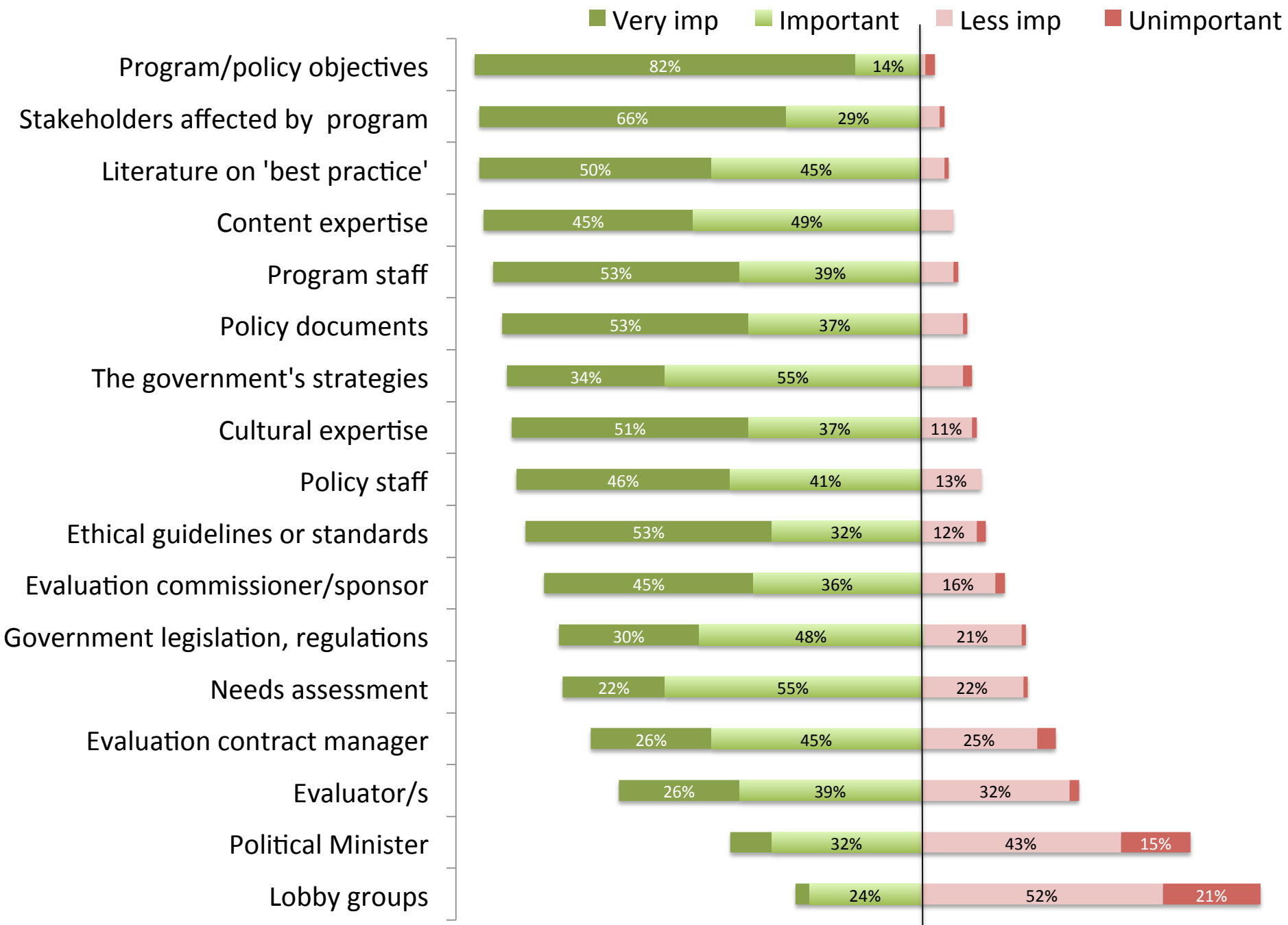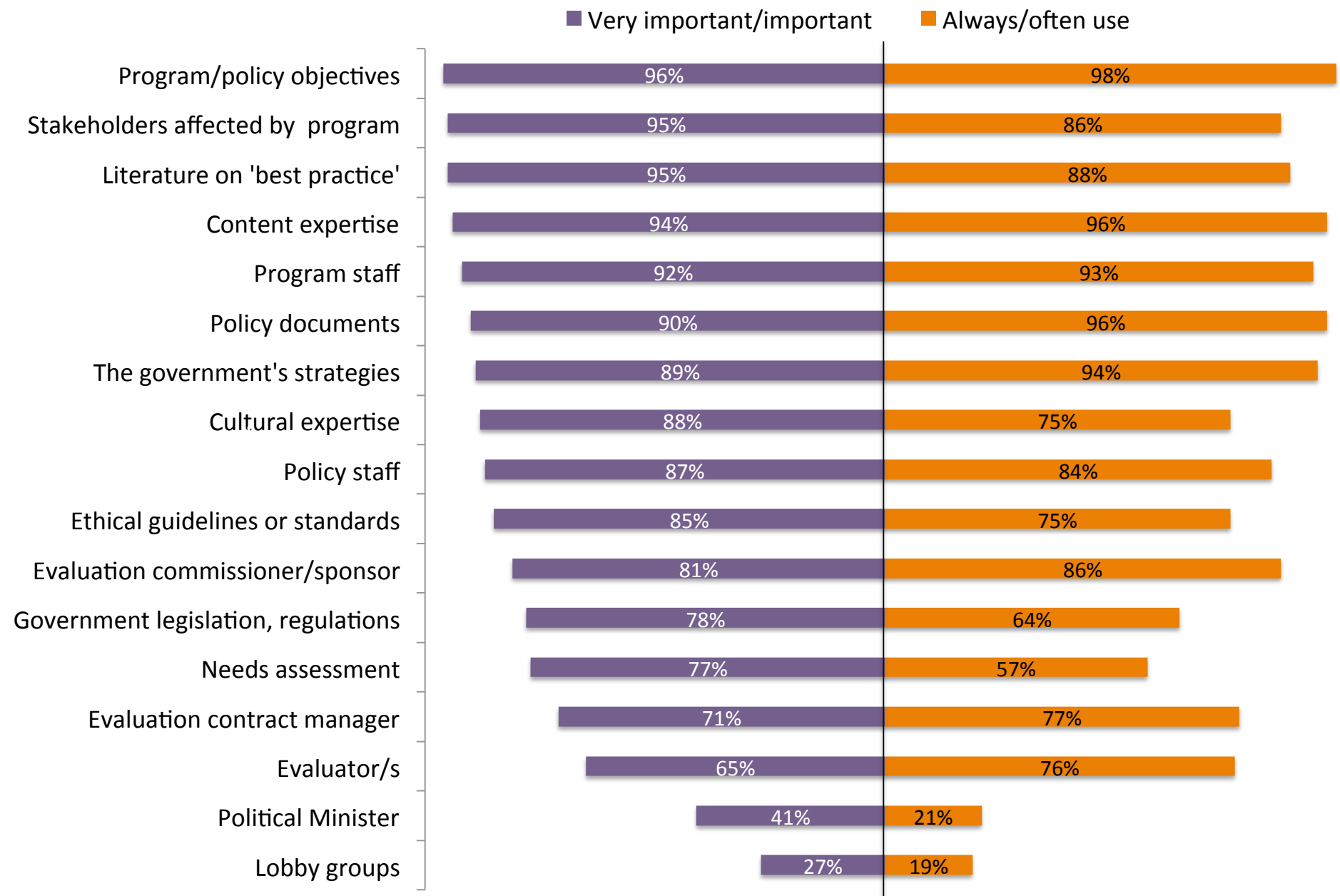|	Very imp	Important	Less imp	Unimportant
Program/policy objectives	82%	14%		
Stakeholders affected by program	66%	29%		
Literature on 'best practice'	50%	45%		
Content expertise	45%	49%		
Program staff	53%	39%		
Policy documents	53%	37%		
The government's strategies	34%	55%		
Cultural expertise	51%	37%	11%	
Policy staff	46%	41%	13%	
Ethical guidelines or standards	53%	32%	12%	
Evaluation commissioner/sponsor	45%	36%	16%	
Government legislation, regulations	30%	48%	21%	
Needs assessment	22%	55%	22%	
Evaluation contract manager	26%	45%	25%	
Evaluator/s	26%	39%	32%	
Political Minister	32%	43%	15%	
Lobby groups	24%	52%	21%	

Comparing importance and use

*The next chart compares importance (purple bars) and use (orange bars). Note that **affected stakeholders** and **literature on best practice** are in the top five sources considered important, but are not in the top five sources most commonly used. Also, check out **needs assessment** and the **political ministe**r: 20% more respondents consider these sources important, but are not using it as a source.*

	Very important/important	Always/often use
Program/policy objectives	96%	98%
Stakeholders affected by program	95%	86%
Literature on 'best practice'	95%	88%
Content expertise	94%	96%
Program staff	92%	93%
Policy documents	90%	96%
The government's strategies	89%	94%
Cultural expertise	88%	75%
Policy staff	87%	84%
Ethical guidelines or standards	85%	75%
Evaluation commissioner/sponsor	81%	86%
Government legislation, regulations	78%	64%
Needs assessment	77%	57%
Evaluation contract manager	71%	77%
Evaluator/s	65%	76%
Political Minister	41%	21%
Lobby groups	27%	19%

What makes a source more important?

Respondents were asked: "In your view, what makes a source more (or less) important when identifying dimensions of quality or value"?

The themes that emerged from the data can be grouped around three inter-related dimensions. First, the interests of different stakeholders, and particularly those who are most vulnerable, need to considered. Second, a source needs to be reliable, validated, and relevant to the context. Finally, sources need to be credible to different audiences.

I'm struck by the similarity between these dimensions, and the criteria of justice, truth and beauty that House has referred to as guiding evaluation quality (House, 2014). On the next page, selected respondents' comments illustrate these dimensions.

The source must:

'bring a consciousness of context, culture & power analysis' *(justice)*

'articulate the underlying need or original intent' *(truth)*

'be credible to end users' *(beauty)*

3 factors influencing the selection of sources

1) **Political imperatives**
2) **Feasibility**

3) **Evaluator perspective**
   Managerial
   Democratic pluralist
   Social equity

*Analysis of the open-ended responses also identified three factors that influence what sources respondents use to develop criteria. The first two are often outside the control of the evaluator. Ability to use a source may depend on the significance of the program to government, or the biases & degree of control of the manager or sponsor within an agency (**political imperatives**). The evaluation budget and timeframe influence what is **feasible** to take into account. A third factor is the **evaluator's perspective** about how an evaluation ought to be conducted. Evaluators may focus primarily on serving the needs of managers or internal stakeholders.*

These evaluators give preference to sources within the agency (e.g. the commissioner). Other evaluators are democratic pluralists, aiming to include all legitimate stakeholder interests (Greene, 1997). Then there are evaluators who bring an explicit equity lens to their work, ensuring a focus that is inclusive of minority groups.

So what?

Should we be concerned that program objectives is so high on our list of 'go to' sources for developing criteria? Not only used, but also considered important. Scriven has argued:

> "Goals have nothing to do with merit, only with management monitoring. Serious program evaluation must dig for the fundamental facts that determine merit – the facts of needs and performance and process – and bypass the bog containing the rhetoric of goals and objectives." *(Scriven, 1991. p179)*

Scriven does concede, in his Key Evaluation Checklist (Scriven 2013), that you can keep the project's goals and objectives in mind, and report on success in achieving them, but my guess is that he considers them much less important than many of the respondents in this survey.

Some interesting questions have surfaced from the analysis, such as:

- *How critically do evaluators look at the value sources they include, or don't include, in an evaluation?*

- *What would evaluation criteria look like if we paid better attention to the characteristics of justice, truth and beauty?*

What now?

This presentation has focused on some of the high level findings of the survey, including what sources respondents are using, which sources they consider more important, and why. I'm interested in your views. Are there any surprises or puzzles that you think I could, or should, explore in more depth? In the next phase of this study, I'm keen to explore the characteristics of justice, truth and beauty as they relate to criteria for public sector evaluations in Australia and New Zealand. Maybe there are other characteristics that need to be explored?

My aim that is in four years time, when I've finished this Ph.D., I am confident about identifying, and am able to include, all the relevant values required for a defensible evaluation. Hopefully there will also be some useful learning to share with evaluators!

My broader vision is that public sector agencies embrace the explicit inclusion of relevant values in all evaluations. That's one big step, I think, toward defensible evaluations that support the development of sustainable and equitable public sector policies and programs.

References

Greene, J.C. (1997). Evaluation as Advocacy. *American Journal of Evaluation,* 18.

Fournier, D.M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions for Evaluation,* 68.

House, E. R. (2014). Origins of the ideas in *Evaluating with Validity*. In J.C. Griffith & B. Montrosse-Moorhead (Eds.), *Revisiting truth, beauty and Justice: Evaluating with validity in the 21st century. New Directions for Evaluation,* 142.

Scriven, M. (1991). *Evaluation Thesaurus* (Fourth ed.). Newbury Park: Sage.

Scriven, M. (2013). *Key Evaluation Checklist.*