



# Extreme Genuine Evaluation Makeovers (XGEMs)



<http://RealEvaluation.com>

Keynote address



September 3, 2010

**E. Jane Davidson, Ph.D.**  
<http://RealEvaluation.com>



## The blog: Genuine Evaluation

### What's in:

1. Right questions
2. Value-based
3. Credibly evidenced
4. Well communicated
5. Usable/actionable
6. Sincere intent to use
7. Humble
8. Value for money

### What's out:

- ❑ Value-free "evaluation"
- ❑ Designs that don't answer the questions
- ❑ Objectives-based only
- ❑ Summaries of opinions
- ❑ Average effects only
- ❑ Misleadingly reported
- ❑ Buried or censored

Patricia Rogers & Jane Davidson's blog:  
<http://GenuineEvaluation.com>



- "The need to demonstrate value for money underpins accountability to Ministers and Parliament, effective management, and the ability of New Zealanders to understand how their taxes are spent."

-- *New Zealand Treasury (2008)*



... but

does anyone ever ask about

value-for-money  
***evaluation?***



## Every year, commissioners are faced with disappointing evaluations

---



## 5 species of waste-of-money evaluation

---





## 1. The “no questions” evaluation

- No questions → no answers!
- What you get:
  - Description of the evaluand
  - Feedback from recipients
  - “Outcomes of interest”
  - “Interesting stories”
  - Areas for improvement
- The end product?



## 2. The “wrong questions” evaluation

- Wrong questions → irrelevant answers
- What you get:
  - Questions that miss the point!
  - Unactionable answers
- The end product? →





### 3. The “no answers” evaluation

- Questions, but no answers??
- What you get:
  - Questions up front
  - Leap straight into data collection
  - No evidence weaving
  - Lost in the details
  - No answers!
- The end product?



### 4. The “descriptive” evaluation

- Descriptive questions → Descriptive answers
- What you get:
  - “What’s so?”
  - Stats, stories, opinions
  - Impressive analyses
  - Master’s thesis layout
  - No evidence weaving
- ... no “So what?”
- The end product:





## 5. The “logical leap” evaluation

□ Evaluative questions → evaluative answers

□ What you get:

- Evaluative questions
- Descriptive evidence
- No clear definition of ‘quality’ or ‘value’
- No evaluative inference you can follow
- Sweeping evaluative conclusions

□ The end product:



## Leaping the critical ‘values’ step





## XGEM-ing “the product”

- The questions need to “get” the purpose and cover the big picture issues
- Evaluative questions are needed, to yield more actionable answers:
  - What’s so? → So what? → Now what?
- Need transparent ways of combining descriptive evidence with ‘values’
- Need reporting that gets to the point!



## Descriptive questions ask “What’s so?”

- Was the initiative implemented according to specifications?
- How many people were reached?
- What was the experience like for participants?
- How did participants rate the initiative?
- What were the effects of the initiative on these outcomes?
- *Has there been a [statistically significant] change in these outcomes?*

## Evaluative questions ask “So what?”



- ▣ How **well designed and implemented** was this initiative?
- ▣ How **substantial** was its contribution to the key outcomes?
- ▣ How **educationally significant** were the student outcomes?
- ▣ Is/was this **more effective** than what we were doing before?
- ▣ Was it **worth the investment** given the outcomes it generated?
- ▣ How **sustainable** is this initiative?

## But how do we answer evaluative questions?



There's ...

- How [many] evaluations “do” answer these ... (2 more species for the list!)
- And how we “could” or “should” (?) answer them ...



## 6. The 'Rorschach inkblot' approach

"You work it out"



## 7. The 'divine judgement' approach



*"I looked  
upon it and  
saw that it  
was good"*

... OR we can actually be systematic and transparent about it



Evaluative rubrics make the 'values' step explicit & transparent



- Evaluative rubrics:
  - A broad-brush way of transparently defining what good, excellent (etc) quality, value or performance would look like in practice
  - Allow interpretation of qualitative, quantitative and mixed method data
    - *as a set*



## Evaluative rubrics define what different levels of performance “look like”

Performance Rating	Performance Descriptors for Answering Key Evaluation Questions
<b>Excellent</b>	Performance is clearly very strong or exemplary in relation to the question. Any gaps or weaknesses are not significant and are managed effectively.
<b>Good</b>	Performance is generally strong in relation to the question. No significant gaps or weaknesses, and less significant gaps or weaknesses are mostly managed effectively.
<b>Adequate</b>	Performance is inconsistent in relation to the question. Some gaps or weaknesses. Meets minimum expectations/ requirements as far as can be determined.
<b>Poor</b>	Performance is unacceptably weak in relation to the question. Does not meet minimum expectations/requirements.
<b>Insufficient evidence</b>	Evidence unavailable or of insufficient quality to determine performance.

Source: NZQA’s External Evaluation & Review framework



## Rubrics demand systematic use of evaluative inference to draw conclusions

e.g. when rating “Good”		you need to show ...
Key points from performance descriptors	Performance is <u>generally strong</u> in relation to the question.	Specifically, what evidence led you to believe performance was “generally strong” – as opposed to “clearly very strong or exemplary” (excellent) or “inconsistent” (adequate)? Include the most important examples of BOTH positive and negative evidence.
	No <u>significant</u> gaps or weaknesses,	What were the gaps or weaknesses, and why should they be considered “not significant”? Based on what?
	and less significant gaps or weaknesses are <u>mostly managed effectively</u> .	What, specifically, is the tertiary education organisation doing to manage gaps and weaknesses, and why do you consider this “effective management” in most or all instances?

## Rubrics may be very criterion-specific, e.g. Parent & whānau engagement in education



Rating	Description
<b>Highly effective</b>	<ul style="list-style-type: none"> <li>□ Parents/whānau are extremely well-informed, confident and highly engaged in their children's education in ways that maximise the children's potential.</li> <li>□ Parent and whānau knowledge and perspectives are well respected, highly valued and fully integrated in ways that benefit the children's education.</li> <li>□ Māori content and language are clearly evident and infused in ways that are appropriate for local whānau.</li> </ul>
<b>Minimally effective</b>	<ul style="list-style-type: none"> <li>□ Levels of parent/whānau/caregiver engagement are <u>just sufficient</u> to support children's education, although there is significant room for improvement</li> <li>□ The school demonstrates understanding of Māori, Pasifika and other cultures, including the concepts of whānau, co-parenting and other family structures.</li> </ul>
<b>Poor or Detrimental</b>	<p><u>Any one or more</u> of the following:</p> <ul style="list-style-type: none"> <li>□ Levels of whānau engagement are <u>extremely low</u> or are <u>deteriorating</u> – to an extent that adversely impacts children's education</li> <li>□ Whānau report being talked "at" or down to, made to feel unwelcome or stupid, or that their perspectives are disrespected or sidelined</li> <li>□ Information is either withheld or presented in ways that prevent meaningful whānau involvement</li> </ul> <p style="text-align: right; font-size: small;">Source: MOE projects (various)</p>

## The 'values' step must be explicit



- In 'divine judgement' and 'logical leap' evaluation, values ARE being applied; they are just not being made explicit
- A transparent 'values' step allows:
  - Genuine evaluative conversations (with more voices at the table) about what 'success' looks like
  - The definitions of 'success' and their application are more easily challenged – and therefore improved

## Structure evaluation reports so they *demand* actionable answers

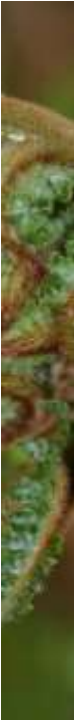


- Executive Summary
  - 2 pages
  - 7 +/- 2 headings
  - Each heading is one of your key evaluation questions
  - 1-2 paragraphs under each heading
  - These 1-2 paragraphs directly and succinctly answer the question!
- Introduction
- Methodology (& why)
- Values & Context
- Key Evaluation Questions
- Findings
  - 7 +/- 2 sections
  - Each section heading is one of your evaluation questions
  - Succinct answer, followed by the evidence

## 9 Golden Rules for commissioning waste-of-money evaluations



1. Use an onerous RFP process and high financial barriers to screen out bidders
2. Keep the available budget top secret
3. Always opt for the evaluator with the lower daily rate
4. Opt for subject matter expertise over evaluation expertise
5. Assume "cultural expertise" is covered off if there's a brown face on the evaluation team



## 9 Golden Rules for commissioning waste-of-money evaluations

---



6. Expect detailed plans and full, accurate costings based on RFP documentation
7. Test the evaluators with “guess the project history” games
8. “Set and forget”
9. Bury and forget about disappointing evaluations – just commission another one!



## Current commissioning no good for attracting/selecting the best evaluators

---



- ▣ Detailed plans and costings written ‘blind’
- ▣ 90% boilerplate
- ▣ Time costs a huge barrier to bidding
- ▣ Not good criteria for evaluating proposals!



## Extreme genuine commissioning makeovers

---



- ▣ EoI process to generate a shortlist
- ▣ Identify the 3-5 sticky issues or challenges that will make or break the evaluation
- ▣ Ask evaluators to describe how they approach such challenges
- ▣ Hear the shortlisted evaluators 'thinking on their feet'
- ▣ Select evaluators based on capabilities to handle the challenges, not written plans



## Current commissioning cumbersome for evaluating emergent or strategic work

---



- ▣ Budgets associated with each project, not with where the questions may lie
- ▣ Focus is on 'new money', not the ongoing big ticket items
- ▣ Cumbersome 'variation' paperwork when things (inevitably!) change
- ▣ Difficult to contract for developmental evaluation



## XGEMs → better value for money

- Improve the product:
  - The right questions
  - Transparent evaluative reasoning
  - Clearly communicated, actionable answers
- Better align commissioning processes and client engagement with needs



## References and resources

- Genuine Evaluation blog (Jane Davidson & Patricia Rogers) <http://GenuineEvaluation.com>
- Jane's website: <http://RealEvaluation.com>
- Davidson, E. J. (2004). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.
- Davidson, E. J. & Martineau, J. M. (2007). Strategic uses of evaluation. In *Leadership development evaluation handbook*. San Francisco, CA: Jossey-Bass

