The Rubric Revolution:

Evaluative blending of mixed method evidence

Demonstration session at

AMERICAN EVALUATION ASSOCIATION

Anaheim, CA
November 2, 2011

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

Do we really 'mix' our methods?

- Mixed methods? … or 'both methods'?

  - Problematic 'mixless' reporting structures
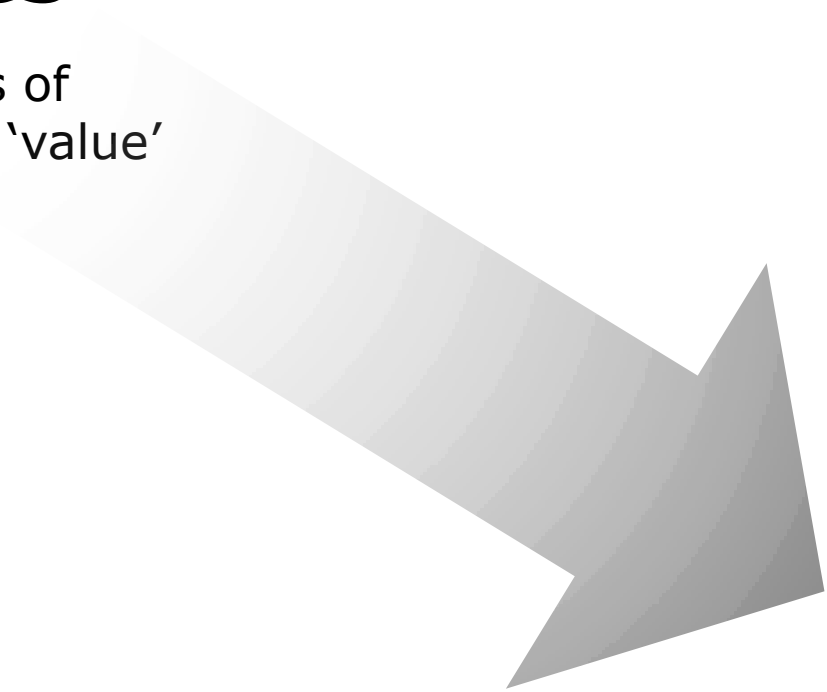
  - Lack of systematic 'mixing' methodologies

Values

Definitions of 'quality' & 'value'

Determining appropriate **criteria** & **evidence**

Values

Definitions of 'quality' & 'value'

("how good is good")

Qualitative & quantitative **evidence**

Evaluative conclusions

(saying something explicit about quality & value)
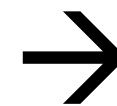
It's what makes evaluation eVALUation!

What's So?   **So What?**

**Descriptive
Facts**
$+$
(qualitative,
quantitative &
mixed method
evidence)

**Definitions of
"quality"
& "value"**

(also factual,
evidence-based
e.g. needs,
potential,
aspirations)

$\rightarrow$

**Evaluative
conclusions**

(saying
something
explicit about
quality, value)

Real evaluation is *one step harder* than descriptive research

□ Descriptive research asks

"What's So?"

■ **What are** the outcomes of a particular intervention? (and why, for whom, etc)

□ Evaluation also asks

"So What?":

■ **How good are** the outcomes …?

■ … and are those outcomes *good enough*?

We need to answer *evaluative* questions

Descriptive questions	Evaluative questions
How was the program implemented?	**How well** (how effectively, efficiently, professionally, culturally appropriately, ethically) was the program implemented?
What were the outcomes?	**How substantial and valuable** were the outcomes?
Did program recipients experience improvements in the key outcomes?	Did the program recipients experience **large enough** improvements in the key outcomes (given the timeframe, investment in the program, and recipients' needs)?
What were the costs?	**How reasonable** were the costs?

Much so-called 'evaluation' skips this "how good is good" step

"*You work it out*"

Others do it with smoke & mirrors ...

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

Rubrics are an attempt to be systematic and transparent about it

What else can we call rubrics?

- **Evaluative interpretation guides** for evidence (qualitative, quantitative, mixed; considered *as a set*)

- **"Ladders of change"**, progressions

- **Definitions of "how good is good"** and "how good is good enough"

- **"Evidence pictures/scenarios"** – which one is the "best fit" with the evidence we have?

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

Rating	Description
Highly effective	◻ Parents/whānau are extremely well-informed, confident and highly engaged in their children's education in ways that maximise the children's potential. ◻ Parent and whānau knowledge and perspectives are well respected, highly valued and fully integrated in ways that benefit the children's education. ◻ Māori content and language are clearly evident and infused in ways that are appropriate for local whānau.
Minimally effective	◻ Levels of parent/whānau/caregiver engagement are <u>just sufficient</u> to support children's education, although there is significant room for improvement ◻ The school demonstrates understanding of Māori, Pasifika and other cultures, including the concepts of whānau, co-parenting and other family structures.
Poor or Detrimental	<u>Any one or more</u> of the following: ◻ Levels of whānau engagement are <u>extremely low</u> or are <u>deteriorating</u> – to an extent that adversely impacts children's education ◻ Whānau report being talked "at" or down to, made to feel unwelcome or stupid, or that their perspectives are disrespected or sidelined ◻ Information is either withheld or presented in ways that prevent meaningful whānau involvement

Sample from 6-level rubric, NZ Ministry of Education

Rubrics can also be generic, to be applied across a range of questions/criteria

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

Source: NZQA's External Evaluation & Review framework

Rubrics demand systematic use of evaluative inference to make ratings

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?

14

Outcome rubrics for a mental health program (from Kate McKegg's work)

A Recovery Matrix

Personal	Clinical
Hope for the future	Daily living skills
Quality of life	Mental Health
Spirituality / Personal Beliefs	Physical Health

| Cultural	Social
Culture	**Housing & accommodation**
Relationships (whānau)	Money and finances
Autonomy	Education, training and work (paid and unpaid

Individual outcome rubric: Housing and accommodation

Crisis	At Risk	Stable / supported	Self-sufficient	Thriving
I am homeless, or about to be evicted: my personal safety is at risk and I am unsafe to remain at home. I am unable to continue living either on my own or with the people I live with. The situation is intolerable	I am at risk of eviction, either because of my actions or because I can not afford the rent. I feel unsafe in my home. My house is in a state of disrepair. I lack some furnishings which I would like. I would like to change my living arrangements.	My housing Is okay – it is warm, dry and affordable. I am safe in my home and have essential furnishings. Some of the time I am ok with who I live with but would like things to be different in the future	I am happy with my housing. My home is comfortable, safe and affordable and furnished to my liking. I am happy with my living arrangements – I like living by myself, with family or housemates.	I am settled and happy in the house of my choice, at this time. It is comfortable, safe and affordable and furnished to my liking. I really enjoy living by myself, with family or with housemates.

16

A graph of change – used for understanding the recovery journey
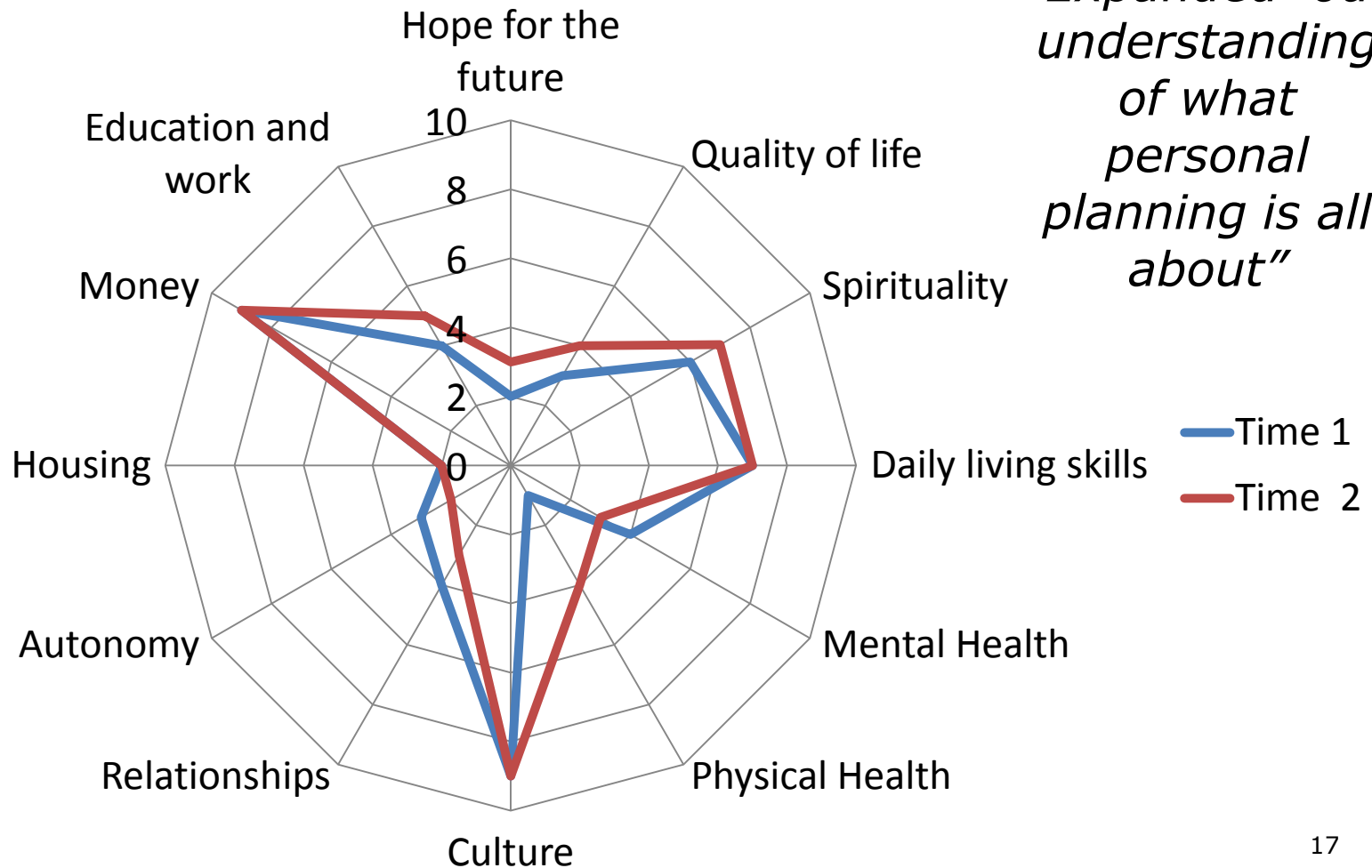
"Expanded our understanding of what personal planning is all about"

How do you create rubrics?

- Get the right people in the room!

- Clearly identify your evaluation question or criterion

- Brainstorm what distinguishes 'highly effective' from 'ineffective' (or worse) performance

- Draw boundaries around what's in and out

- Facilitate the rubric writing process

- Debate; recalibrate; field test; hone

# Using rubrics for strategic, whole-portfolio evaluation

- Ever wondered how to make sense of findings (and synthesize, compare) across a range of projects or programs?

- What if they have been (or, must be) monitored or evaluated using very different methods, measures, and approaches?

- Example: Milestone (monitoring) reports for professional learning and development delivered to schools (teachers and leaders)

PLD Milestone Reports ... before

PLD Providers ...

"Phonebook milestone reports"
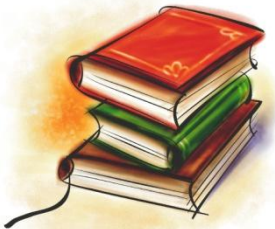
Different focus areas

Different contexts

Different student populations

Different process & outcome criteria

Different methods & evidence

The Client ...

overwhelmed

Unable to make sense across reports [20]

PLD Milestone Reports ... after

The client provides:

A clear set of 6 evaluation questions to be answered in <u>all</u> reports

Set of 6 rubrics to guide how evidence should be interpreted

PLD provider training

PLD Providers:

Reporting covers the same 6 KEQs; the same rubrics are used

Different evidence is used, as appropriate for context & populations

PLD providers do their own self-assessment

More comparable, slimmer reports

And then ...

The client considers self-assessments, provides initial feedback

Both parties engage in an *evaluative conversation* to explore differences

Client make sense of findings across multiple projects

# Aren't rubrics all just subjective?

Important to understand the three kinds of subjectivity:

1. Arbitrary, idiosyncratic, unreliable, and/or highly personal (i.e., based on personal preferences and/or cultural biases)

2. Assessment or interpretation by a person, rather than a machine or measurement device, of something external to that person (e.g., expert judgment of others' skills or performance)

3. About a person's inner life or experiences (e.g., headaches, fears, beliefs, emotions, stress levels, aspirations), all absolutely real but not usually independently verifiable

Plus the red herring: Subjective vs. objective measures

Why rubrics?
Why not just indicators?

Indicators

Outcome domain

Criterion
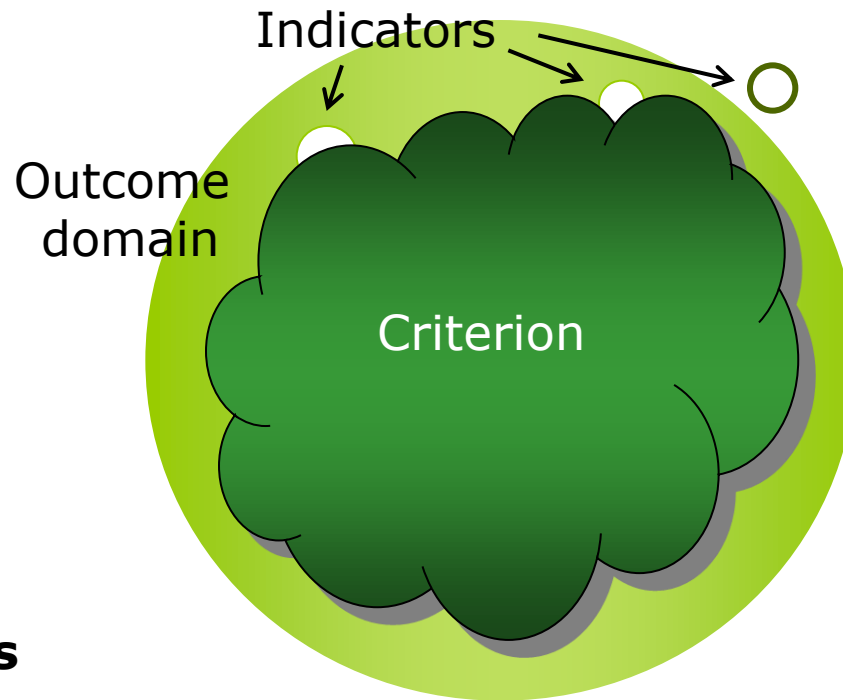
Indicators

True Criteria

Easy to measure
Precise
Narrow
Manipulable

Harder to measure
Approximate
Broad-brush
Unmanipulable

The value of rubrics

- Serious about values - perfect for ensuring community, cultural, & organizational values/aspirations incorporated

- Identify multiple levels of performance and progress, not just one vaguely 'acceptable' level

- Cover the vast majority of the process or outcome domain => largely unmanipulable

# The value of rubrics

◻ Versatile – they work well for:
  - participatory or 'independent' evaluation
  - "fairly quick and fairly clean" or in-depth analysis
  - qualitative, quantitative – but especially good for mixed methods
  - needs assessment (baseline) and outcomes
  - overviews and drill down

◻ Build shared understanding of what constitutes effectiveness – provided you get the 'languaging' right!

References and resources

- Innovative uses of rubrics in tomorrow morning's Presidential Strand session (Thursday 8am in Pacific A): *The Agent, the Provocateur, the Activist: Creative evaluators, surfacing undercover values in a rubric revolution*

- Davidson, E. J. (2004). *Evaluation methodology basics: The nuts and bolts of sound evaluation.* Thousand Oaks, CA: Sage.

- Davidson, E. J. (2007). Unlearning Some of our Social Scientist Habits. *Journal of MultiDisciplinary Evaluation, 4*(8), iii-vi. http://tiny.cc/unlearning

- Genuine Evaluation blog (Jane Davidson & Patricia Rogers) http://GenuineEvaluation.com

- Jane's website: http://RealEvaluation.com

- Kinnect Group website http://kinnect.co.nz/
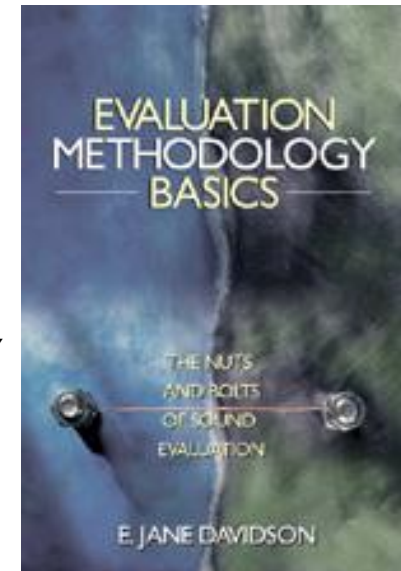
Examples of Evaluation Rubrics

- Self-review tool for schools: Focus on students achieving below curriculum expectations in literacy. http://tiny.cc/literacytool

- Davidson, E.J., & Wehipeihana, N. (2011) 'Credible evidence of effectiveness for Māori learners: An introduction to the Measurable Gains Framework & related rubrics' presented at the National Aspriring Principals Programme Hui, 18 April, Auckland.
  http://kinnect.co.nz/wp-content/uploads/2011/08/110418-Davidson-Wehipeihana-NAPP.pdf

- Rubrics for gauging progress against *Ka Hikitia*, the NZ Ministry of Education's Maori education strategy: http://tiny.cc/kahikitia

- NZQA's External Evaluation & Review framework for evaluative quality assurance of tertiary (higher) education organizations: http://www.nzqa.govt.nz/providers-partners/quality-assurance-of-itps/external-evaluation-review/

- Rubrics used in a mental health setting (from Kate McKegg): http://www.wellink.org.nz/pdfs/CharacteristicsofGoodPeerSupport.pdf