The Evaluator’s Methodological Toolkit

A Helicopter Tour with a Few Landings

E. Jane Davidson, Ph.D., 1-30-03
The Evaluation Center & Department of Sociology
http://homepages.wmich.edu/~j davidso

Tour Itinerary

- What is evaluation?
- What’s in a methodological toolkit ...
  - For applied social scientists?
  - For evaluators?
- A closer look at two methodologies:
  - Importance determination
  - Merit determination
- Where to next with development of the evaluator’s methodological toolkit?
What is evaluation, exactly?

- Evaluation is the systematic determination of the quality, value, or importance of something.

- e-VALU-ation has two components:
  - Descriptive + Values = Evaluative
  - Facts + (what’s so) = Conclusions (so what)

- Values are what get us from “What’s so?” to “So what?”
  - and they are the trickiest part of an evaluation!

What is evaluation for?

- Good evaluation can provide useful information for strategic decision making.

Innovation → Evaluation → Decisions

Try something new...

- Is it better than what we had before?
- Is it better than the alternatives?
- How might it be improved?
- Is it worth the additional cost?

Improvement/streamlining

Choices among alternative options
Evaluation: The basic steps

- Identify stakeholders & their information needs
- Identify key evaluation questions
- Identify all the evaluative criteria
  - What distinguishes a good one of these from a poor one?
- Determine importance of criteria
- Gather information/data
- Determine merit
  - How good is OK/good/excellent?
  - How poor is unacceptable?
- Identify useful comparisons
- Synthesize all of the above → answer the evaluation questions

The social scientist’s approach

- When designing and conducting applied research:
  - How do you define research questions (e.g., choose a thesis/dissertation topic)?
  - How do you identify variables to include in your study?
  - What tools or methods are available for analyzing your data?
What else would an evaluator need?

- Think of an intervention [X] from your field ...
- What tools/strategies are available for:
  - Identifying all of the criteria that distinguish a good (effective, valuable, etc) [X] from a poor one?
  - Figuring out which of these criteria are extremely important vs. important vs. marginally important?
  - Defining what level of performance on each criterion constitutes “excellent” vs. “OK” vs. “unacceptable”?
  - Synthesizing mixed results (pros and cons) on 20-30 variables → conclusion about X (your evaluand)?

A hypothetical example

<table>
<thead>
<tr>
<th>Criterion #1</th>
<th>Criterion #2</th>
<th>Criterion #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Performance -------------->
The importance of importance

Criterion #1: Extremely important
Criterion #2: Moderately important
Criterion #3: Of minor importance

Performance -->

Importance Determination

- Use expert judgment
- Have stakeholders “vote”
  - Whose views should you poll?
  - How should you ask them? *
- Investigate for yourself:
  - Probing interviews of key informants \(\rightarrow\) judgment *
  - Evidence from the literature
  - Causal linkages *
Having stakeholders “vote”

- Survey
  - Rate importance; average responses
  - Paired comparisons → ranked importance
- Group process
  - Rate importance, pool ratings, discuss differences, negotiate agreement
- Key issues:
  - When would opinion NOT be the right way to determine importance?
  - Who should you ask/NOT ask for input?

Probing interviews + judgment

**Focus:** Which outcome criteria are the most important here?

**Probe:** Actual or probable impact:

- How beneficial would it be overall if performance on this criterion were really excellent?
- How detrimental would it be overall if performance on this criterion were really poor?
- How do you know? Based on what evidence/experience?
Ex: The “Organizational IQ Test”

**CULTURE**
- High-Performance Culture
- Challenging Assumptions
- Shared Vision & Intuition
- Team Learning & Communication
- Systems & Nonlinear Thinking
- External & Future Scanning
- Innovation & Experimentation
- Systematic Evaluation

**OUTCOMES**
- Feeling able to add max. value
  - *extremely important*
- Retention of top talent
  - *v. important*

**Add strengths of links**

**CULTURE**
- High-Perf’ce Culture
- Challenging Assmpt’ns
- Shared Vision/Intuition
- Team Learning/Comm’n
- Systems Thinking
- Ext/Future Scanning
- Innov’n & Experiment’n
- Systematic Evaluation

**OUTCOMES**
- Feeling able to add max. value
  - *ex. important*
- Retention of top talent
  - *v. important*
Combining the info

**CULTURE**
- High-Perf’ce Culture
- Challenging Assmptns
- Shared Vision/Intuition
- Team Learning/Comm
- Systems Thinking
- Ext/Future Scanning
- Experimentation
- Systematic Evaluation

**OUTCOMES**
- Feeling able to add max. value
  - *ex. important*
- Retention of top talent
  - *v. important*

---

**Merit Determination Options**
- Evaluate relative to **goals** (weak option)
- Have stakeholders “vote” on standards
- Evaluate individual indicators relative to some standards (e.g., fixed **quality standards**)
- **Benchmark** against other efforts, alternatives, or competitors
- **Interview/probe** real impact on recipients needs: How substantial is this impact in people’s lives?
- Use **rubrics** that combine several of the above
### A basic rubric for determining merit

<table>
<thead>
<tr>
<th><strong>Excellent</strong></th>
<th>Clear example of best practice in this domain; no weaknesses.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Very Good</strong></td>
<td>Very good or excellent performance on virtually all aspects; no real weaknesses.</td>
</tr>
<tr>
<td><strong>Good</strong></td>
<td>Reasonably good performance overall; may have a few slight weaknesses.</td>
</tr>
<tr>
<td><strong>Satisfactory</strong></td>
<td>Fair performance - some serious (but non-fatal) weaknesses on a few aspects.</td>
</tr>
<tr>
<td><strong>Poor</strong></td>
<td>Clear evidence of impaired functioning on this dimension; very serious weaknesses.</td>
</tr>
</tbody>
</table>

---

### Ex: High-Performance Culture – V.G.

- **Striving for excellence** – **excellent**
  - Strong consensus of high-performance culture
  - Median score 4.5/5.0

- **Seeking out criticism** – **very good**
  - Owner/manager excellent role model for this, though not all employees aware how open he is to criticism
  - Median scale score 3.7/5.0

- **Understanding the ‘performance gap’** – **very good**
  - Ambiguity in innovative work prevents consensus on this
  - Median scale score 4.0/5.0
The Learning Culture ‘Snapshot’

- Innovation & Experimentation
- Systematic Evaluation
- Challenging Assumptions
- Shared Vision & Intuition
- External & Future Scanning
- High-Performance Culture
- Team Learning/Communication
- Systems & Nonlinear Thinking

The Methodological Toolkit

- Identification of all relevant criteria
  - Multilevel needs assessment
  - Search for other relevant values
- Determining importance of criteria
- Determining merit of performance on criteria
- Synthesis methodology
  - Pulling together the pros and cons to draw overall conclusions
Where to next?

- Not reinventing the wheel:
  - What other disciplines are grappling with the same problems?
  - What have they developed?
- Exporting proto-methodologies
  - How can evaluation-specific methodologies be applied in a variety of fields/settings?
  - What can we learn from critiques of how well it works (validity, simplicity, relevance, utility)?
- Interdisciplinary brainstorming!