Using qualitative and mixed methods in project evaluation

Michael Woolcock
Lead Social Development Specialist
Development Research Group, World Bank

Paris
13 November 2012
Overview

• From projects to questions to methods to data
  – ‘Rigor’ not a feature of a particular method
• Four central tasks in evaluation:
  – Construct, internal and external validity...
  – ...incorporating an explicit theory of change
• The value and ‘value-added’ of qualitative approaches
• Options, strategies for assessing projects using qualitative and mixed methods
• Examples of using qual and ‘mixed’ methods
  • See [www.sagepub.com/rwe](http://www.sagepub.com/rwe)
• Barron, Patrick, Rachael Diprose and Michael Woolcock (2011) *Contesting Development: Participatory Projects and Local Conflict Dynamics in Indonesia* New Haven: Yale University Press
On elections, and evaluation...

“[T]he real lesson from the accuracy of Silver’s predictions [of the US elections is] not that numbers beat words, not that quantitative research is inherently superior to qualitative investigation, but that whatever type of knowledge and information we deal with, not least professionally, we need to critically examine its empirical and epistemological premises.”

-- Cornel Sandvoss, Univ. of Wisconsin

http://blog.commarts.wisc.edu/2012/11/08/methods-of-failure-how-political-journalism-lost-the-us-presidential-election-to-nate-silver/comment-page-1/#comment-370787
Types of Projects

Four *analytical* questions

<table>
<thead>
<tr>
<th>Is your activity...</th>
<th>That is, does the success of your policy require...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transaction Intensive?</td>
<td>Many agents to act or few, over extended time periods?</td>
</tr>
<tr>
<td>Locally Discretionary?</td>
<td>Implementing agents to make finely based distinctions about the “state of the world”? Are these distinctions difficult for a third party to assess?</td>
</tr>
<tr>
<td>Based on Known Technology?</td>
<td>That agents innovate to achieve desired outcomes?</td>
</tr>
<tr>
<td>High Stakes?</td>
<td>Agents to resist large temptations to do something besides implement the policy?</td>
</tr>
</tbody>
</table>
## Classification of “activities” in health

<table>
<thead>
<tr>
<th>Activity</th>
<th>Locally Discretionary</th>
<th>Transaction Intensive</th>
<th>High Stakes Known ‘Technology’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iodization of salt</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Technocratic (implementation light; policy decree)</td>
</tr>
<tr>
<td>Vaccinations</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistical (implementation intensive, but easy)</td>
<td></td>
</tr>
<tr>
<td>Ambulatory curative care</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Implementation Intensive ‘Downhill’ (of services)</td>
<td></td>
</tr>
<tr>
<td>Regulation of private providers</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Implementation Intensive ‘Uphill’ (of obligations)</td>
<td></td>
</tr>
<tr>
<td>Encouraging preventive health</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complex (implementation intensive, motivation hard), need (continuous?) innovation</td>
<td></td>
</tr>
</tbody>
</table>
Arraying methods in social science

Breadth

Household surveys; Census

Coded textual analysis (e.g., of newspapers)

Targeted key-informant surveys

Case studies, Interviews, Observations
Contesting Development
Participatory Projects and Local Conflict Dynamics in Indonesia

PATRICK BARRON
RACHAEL DIPROSE
MICHAEL WOOLCOCK

Yale University Press, 2011
Strengths of qualitative methods

• Dynamics, not just demographics, of identity
• Accessing marginalized, stigmatized groups
• Unpacking context idiosyncrasies
• Building rapport, credibility with clients
• Engaging alternative approaches to causality
  – Econometrics vs. law vs. history vs anthropology
• Improving quantitative approaches
  – Observing ‘unobservables’; enhancing construct validity
• Exploring characteristics of (apparent) outliers
• Helping resolve anomalies (methods, people)

➢ Strengths and weakness of quant and qual are complements, not substitutes
Making, assessing impact claims

Quality of empirical knowledge claims turns on...

1. **Construct validity**
   - Do key concepts (‘property rights’, ‘informal’) mean the same thing to different people? What gets “lost in translation”?

2. **Internal validity**...
   - In connecting ‘cause’ (better schools) and ‘effect’ (smarter children), have we considered other factors that might actually be driving the result (home environment, community safety, cultural norms)? Programs rarely placed randomly...

3. ...assessed against **a ‘theory of change’**
   - Specification of how project’s components (and their interaction) and processes generate outcomes
   - Impact trajectory over time: where by when?

4. **External validity** (how generalizable are the claims?)
   - If it works here, will it work there? If it works with this group, will it work with that group? Will bigger be better?
1. Construct validity

• Examples of how mixed methods in action:
  – Serious field testing of questionnaire items, and their sequencing
    • NOT cut-and-paste from elsewhere
  – ‘Anchoring vignettes’ (Gary King et al)
    • Assessing “quality of government” in China and Mexico
2. Internal validity

In Evaluation 101, we assume...

\[ \text{Impact} = f(\text{Design}) \mid \text{Selection, Confounding Variables} \]

Adequate for ‘simple’ interventions with a ‘good-enough’ counterfactual.

But this is inadequate for assessing ‘complex’ interventions:
* design is multi-faceted (i.e., has high ‘causal density’)
* interaction with context is pervasive, desirable
* implementation quality is vital
* trajectories of change are probably non-linear (perhaps unknowable ex ante)
Evaluating ‘complex’ projects

Impact = f ([Design quality, Causal density], Implementation, Context) | Selection effects, Confounding variables, Reasoned expectations

In ‘complex’ projects such as those shaping the business environment:
* Causal density is high, loose, often unseen
* Implementation and Contexts are highly variable
* Reasoned expectations are often unknown (perhaps inherently unknowable ex ante)
Pervasive problem

• Such projects are inherently very complex, thus:
  – Very hard to isolate ‘true’ impact
  – Very hard to make claims about likely impact elsewhere
  – Understanding how (not just whether) impact is achieved is also very important
    • Process Evaluations, or ‘Realist Evaluations’, can be most helpful (see work of Ray Pawson, Patricia Rogers et al)
    • Mixed methods, theory, and experience all crucial for investigating these aspects
3. Assessing impact against a theory of change: Understanding impact trajectories
Understanding impact trajectories

"Same" impact claim, but entirely a function of *when* the assessment was done...
Understanding impact trajectories

If an evaluation was done at ‘A’ or ‘B’, what claims about impact would be made?
Understanding impact trajectories
4. External validity

• Logic of (elite) research and most ‘development effectiveness’ debates leads to a focus on Design
  – The better to identify, replicate ‘best practices’, ‘tools’
  – Preferably validated via an RCT (the ‘gold standard’)

4. External validity

• Logic of (elite) research and most ‘development effectiveness’ debates leads to a focus on Design
  – The better to identify, replicate ‘best practices’, ‘tools’
  – Preferably validated via an RCT (the ‘gold standard’)
  – (RCTs fine for ‘simple’ projects, or aspects thereof)

• To better address IV and EV, especially of complex interventions, we need theory and an array of methods to best match the type of project
  – The better to specify the conditions under which certain outcomes are likely to be observed
  – Enhancing frequency and rigor of case studies is crucial
Putting it all together

<table>
<thead>
<tr>
<th>Project Design Features</th>
<th>Technocratic</th>
<th>Logistical</th>
<th>Implementation Intensive (‘Downhill’)</th>
<th>Implementation Intensive (‘Uphill’)</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation Quality</td>
<td>Strong</td>
<td>Weak</td>
<td>Strong</td>
<td>Weak</td>
<td></td>
</tr>
<tr>
<td>Context Compatibility</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Even with low EV interventions, the **ideas** and **processes** behind them may still travel well.
Putting it all together

<table>
<thead>
<tr>
<th>Project Design Features</th>
<th>Technocratic</th>
<th>Logistical</th>
<th>Implementation Intensive (‘Downstream’)</th>
<th>Implementation Intensive (‘Upstream’)</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation Quality</td>
<td>Strong</td>
<td>Weak</td>
<td>Strong</td>
<td>Weak</td>
<td>Strong</td>
</tr>
<tr>
<td>Context Compatibility</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Even with low EV interventions, the **ideas** and **processes** behind them may still travel well.
Putting it all together

<table>
<thead>
<tr>
<th>Project Design Features</th>
<th>Technocratic</th>
<th>Logistical</th>
<th>Implementation Intensive ('Downstream')</th>
<th>Implementation Intensive ('Upstream')</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation Quality</td>
<td>Strong</td>
<td>Weak</td>
<td>Strong</td>
<td>Weak</td>
<td></td>
</tr>
<tr>
<td>Context Compatibility</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Even with low EV interventions, the **ideas** and **processes** behind them may still travel well.
Practical examples (1)

1. Effects of ethnicity on poverty in Guatemala
   – ‘Parallel’, part of a large mixed-methods assessment
   – **Quan**: expanded national household survey
     • first social capital module
     • large differences by region, gender, income, ethnicity
     • pervasive elite capture
   – **Qual**: 10 villages (5 different ethnic groups)
     • perceptions of exclusion, access to services
     • fear of reprisal, of children being stolen
     • legacy of shocks (political and natural)
     • links to LSMS data
Practical examples (2)

2. Everyday governance mechanisms and poverty in Delhi slums (Jha, Rao and Woolcock 2007)
   – ‘Sequential’, also part of large, mixed methods study
   – **Qual**: 4 migrant communities
     • near, far, recent, long-term
   – **Quan**: 800 randomly selected representative households
   – From survival to mobility
     • role of norms (sharing, status) and networks (kinship, politics)
     • housing, employment transitions
     • property rights
   – Understanding ‘governance’
     • managing collective action
     • crucial role of service provision
Concluding thoughts

• Mixed methods as complements, not substitutes
  – All methods have strengths and weaknesses
• The virtues and limits of formal ‘measurement’
  – Tension between simplifying versus complicating reality
  – “Better to be vaguely right than precisely wrong” (Keynes)
• Triangulation
  – Integrating more data, better data, more diverse data
  – Especially important for assessing ‘complex’ interventions
• MM can also be used in rapid assessments, monitoring
  – Similar principles whether large/small, rapid/slow
  – One size (literally) does not fit all
  – M&E fundamentally about learning, adapting