gLOCAL EVALUATION WEEK
Sharing local and global M&E knowledge

Share your experiences. Make new connections.

JUNE 3–7, 2019
Learn more! Go to theclearinitiative.org

CONVENED BY:

IEG | WORLD BANK GROUP
Independent Evaluation Group
World Bank • IFC • MIGA
The unfinished evidence revolution

Howard White
Campbell Collaboration
@campbellreviews @HowardNWhite

June 3, 2019
First wave: the results agenda

1990s

Results agenda
(Outcome monitoring)

Introduced focus on outcomes rather than inputs

Issues

But ‘results frameworks’ don’t tell us what difference the intervention makes...
data are not evidence

Data quality... using validated measures, standardized indicators from indicator pyramid

Indicator proliferation... again standardized indicators
PLUS .....user engagement in setting indicators across the causal chain
MESSAGE FROM THE DUKE OF WELLINGTON TO THE BRITISH FOREIGN OFFICE IN LONDON -- written from Central Spain, August 1812

Gentlemen,

Whilst marching from Portugal to a position which commands the approach to Madrid and the French forces, my officers have been diligently complying with your requests which have been sent by H.M. ship from London to Lisbon and thence by dispatch to our headquarters. We have enumerated our saddles, bridles, tents and tent poles, and all manner of sundry items for which His Majesty's Government holds me accountable. I have dispatched reports on the character, wit, and spleen of every officer. Each item and every farthing has been accounted for, with two regrettable exceptions for which I beg your indulgence. Unfortunately the sum of one shilling and ninepence remains unaccounted for in one infantry battalion's petty cash and there has been a hideous confusion as the number of jars of raspberry jam issued to one cavalry regiment during a sandstorm in western Spain. This reprehensible carelessness may be related to the pressure of circumstance, since we are at war with France, a fact which may come as a bit of a surprise to you gentlemen in Whitehall. This brings me to my present purpose, which is to request elucidation of my instructions from His Majesty's Government so that I may better understand why I am dragging an army over these barren plains. I construe that perforce it must be one of two alternative duties, as given below. I shall pursue either one with the best of my ability, but I cannot do both: 1. To train an army of uniformed British clerks in Spain for the benefit of the accountants and copy-boys in London or perchance, 2. To see to it that the forces of Napoleon are driven out of Spain.

Your most obedient servant Wellington
Questions

Global alignment?

Common metrics?

Who is using for what? (Lesson learning vs accountability)
First wave: the results agenda

1990s

Results agenda (Outcome monitoring)

- Introduced focus on outcomes rather than inputs

Issues

- But ‘results frameworks’ don’t tell us what difference the intervention makes… data are not evidence
- Data quality… using validated measures, standardized indicators from indicator pyramid
- Indicator proliferation… user engagement in setting indicators across the causal chain

Implications for IEG

- Periodic in-depth IEG studies of the state of monitoring in Bank projects (i.e. continue ROSES)
- IEG report on use of “results” by different levels in the Bank
- IEG report on Bank support to data collection and data initiatives (one done, so next one with different focus)
- ICR Reviews already assess the quality of M&E of the project. Identify some issues: (i) inappropriate, vague indicators; (ii) missing indicators; (iii) lack of user engagement (but many cases of effective M&E)

TELL OTHER AGENCIES WHAT YOU ARE DOING AS THEY SHOULD BE TOO!
Monitoring matters

IEG’s work on operational monitoring is important and deserves more attention
Second wave: the rise of impact evaluation

1990s
Results agenda (Outcome monitoring)

2000s Rise of impact evaluation

Issues

- Researcher capture
- Focus on first generation questions
- Partial equilibrium
- Not suitable for all interventions

Is this at all relevant to IEG?
Second wave: the rise of impact evaluation

1990s
Results agenda (Outcome monitoring)

2000s
Rise of impact evaluation

Issues
- Researcher capture
- Focus on first generation questions
- Partial equilibrium
- Not suitable for all interventions

Implications for IEG
- Explicit attention to methods (not only causality)
- Dealing with complexity: address biases
- Unit of analysis in IEG studies: projects or interventions?
<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Theory of change</th>
<th>Case studies</th>
<th>Portfolio review and analysis</th>
<th>Structured literature review / desk review</th>
<th>Semi-structured interviews / interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOE reform (Dec 2018)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Carbon finance (June 2017)</td>
<td>Yes</td>
<td>Yes, will explore possible process tracing or QCA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mobile metropolises (May 2016)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Shared prosperity (April 2016)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Capital Market Development (March 2015)</td>
<td>Yes will develop sector specific indicators of success</td>
<td>Yes. Inferences that are causal will not be made,</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Rural poverty (December 2015)</td>
<td>Yes</td>
<td>Yes using contribution analysis</td>
<td>Yes - detailed description of search in approach paper</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

These are approaches to data collection NOT methods.
Second wave take away point

Pay explicit attention to methods

IEG to better develop methods standards for different types of evaluation question (ongoing)
Wave three: systematic reviews

1990s
Results agenda (Outcome monitoring)

2000s
Rise of impact evaluation
Rise of systematic reviews

Issues
Range of issues and methods
Over use of ‘systematic’
Challenges in qualitative synthesis
Only 30-50% of research can be replicated

Coding error rates are typically 30-50%

Publication bias and outcome reporting bias are prevalent in all literatures
Wave three: systematic reviews

1990s

Results agenda (Outcome monitoring)

2000s  
Rise of impact evaluation

Issues

- Range of issues and methods
- Over use of ‘systematic’

Implications for IEG

- More explicit adoption of SR approaches; e.g.:
  - PICOS
  - Ex ante specification and testing of instruments and methods
  - Double screening and coding
  - Systematic reporting
Research is plagued by high rates of error and bias

Use systematic review methodologies, such as PICOS, ex ante design and testing of coding forms and double coding, which are designed to minimize these biases (IEG evaluations are enormously complex, even interventions being evaluated are slippery)
Fourth wave: knowledge brokering

1990s

Results agenda (Outcome monitoring)

2000s

Rise of impact evaluation

Rise of systematic reviews

Knowledge brokering

Issues

Can over simplify evidence

Need to be context specific

Implications for IEG

More diverse knowledge products

Channels for engagement in strategy development
Knowledge brokering platform pyramid

Checklists

Guidance & guidelines

Evidence portals

Evidence maps

Databases

Systematic reviews

Studies

Data

More heavily brokered knowledge
Choice of intervention for those with severe aortic stenosis

**Transfemoral TAVI**
Inserting a new valve into the aortic valve’s place without open heart surgery. Delivery is through the femoral artery.

**SAVR**
Open-heart surgery, to remove the narrowed aortic valve. Replacement with tissue valve.

### Recommendations

<table>
<thead>
<tr>
<th>Population</th>
<th>Favours TAVI</th>
<th>Favours SAVR</th>
<th>Why?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 85+</td>
<td>Strong</td>
<td></td>
<td><img src="https://via.placeholder.com/50" alt="Why?" /></td>
</tr>
<tr>
<td>Age 75–84</td>
<td>Weak</td>
<td></td>
<td><img src="https://via.placeholder.com/50" alt="Why?" /></td>
</tr>
<tr>
<td>Age 65–74</td>
<td>Weak</td>
<td></td>
<td><img src="https://via.placeholder.com/50" alt="Why?" /></td>
</tr>
<tr>
<td>Age under 65</td>
<td>Strong</td>
<td></td>
<td><img src="https://via.placeholder.com/50" alt="Why?" /></td>
</tr>
</tbody>
</table>

### Key Uncertainties

The major uncertainty is the durability of TAVI valves which drives recommendations in favour of SAVR in younger patients.
Knowledge brokering is an activity in its own right and is much more than traditional dissemination: Innovative interactive ways of presenting evidence are driving evidence use.
Big data, AI and machine learning: a fifth wave?

• Great opportunities for automating aspects of evaluation process (with fewer errors)
• Big data opportunities as Bank data can be linked to other data sources (more in depth analysis of projects)
• Should experiment but with caution and fail-safe.
Overall message

• The work of IEG remains at the forefront of evaluation
• But the investment in methods (and related evidence standards) needs to continue
Thank You

Visit www.campbellcollaboration.org

Join us at wwgs2019.org