SKILLFUL EVALUATION

The Rough Guide to Taking CAIR

SKILLFUL EVALUATION IS …

Competent: The full repertoire of evaluation competencies are available

Astute: The optimal evaluation choices for the evaluation context are deliberatively (intuitively) made

Informed: The evaluation is supported by thorough knowledge of the subject matter

Responsive: The implementation of the evaluation can be adjusted in response to emerging factors

WHAT IS EVALUATION FOR?

The problem of evaluation presenting itself as social science – Prof Linda Richter

Evaluation is a practice that assesses the worth of something by applying rules of evidence, including those of (social) science. – a highly contestable but also defensible description of evaluation.

What is evaluation for?

1. Accountability
2. Demonstrating efficacy of a particular intervention
3. Other ends

When is it for testing the general laws of development intervention/ human behavior/society? ALMOST NEVER

WHAT IS METHODOLOGY?

How do we integrate the following elements to maximise the reliability and validity of findings:

- Primary Elements
  - The *purpose* of the evaluation
  - The *questions* we need to answer
- Secondary Elements
  - The *type of data* we need to collect
  - How we will *collect* it
  - How we will *analyse* it

BROAD METHODOLOGICAL CATEGORIES

Accountability	Efficacy	Other Ends
Economic: *Designs that focus on return on expenditure questions using financial and economic or econometric methods*	Scientific-experimental: *Randomised control trials, quasi-experiments and statistically oriented designs*	Participant oriented: *Designs that emphasise the empowerment of intervention recipients to act*
Audit-managerial: *The collection of data to determine compliance against standards*	Interpretive-ethnographic: *Integration of observational and other qualitative data in a thick description that emphasises meaning in context*	
Rapid: *Design in which the constraints are such that the evaluation depends heavily on expert*	Theory-Driven: *Beginning from an*	

HOW DO WE CHOSE?

To make good choices we consider:

- The purpose of the evaluation, at a utility and a practical level
- The constraints under which the evaluation is to be conducted
- The required burden of proof

PURPOSE

The <u>utility level</u>: the balance between accountability, demonstrating efficacy, and other ends

The <u>practical level</u>: We are evaluating some "thing", what do we need to know about it (what is the primary question being posed?)

1. What is the right thing to do? *Design Evaluation*
2. Are we doing the right thing? *Process Evaluation*
3. Is this thing achieving results? *Outcomes/Impact Evaluation*
4. Why is this thing getting the results its getting? *Diagnostic (and Impact) Evaluation*
5. What have we learned from doing this thing and others like it? *Synthesis Evaluation*
6. What are we getting for what we are spending? *Economic Evaluation*

SPECTRUM OF CAUSALITY

Descriptive		Causal

DESIGN	PROCESS	DIAGNOSTIC	SYNTHESIS	ECONOMIC	IMPACT
The more the findings make causal claims, the higher the burden of proof required.					

CONSTRAINTS

Evaluation related
- Time
- Budget

Implementation Context
- Simple, Complicated and Complex programmes
- Simple, Complicated and Complex systems
- Programme and systems are equally causal

REQUIRED BURDEN OF PROOF

Reliability

Validity

Representivity and sampling

RELIABILITY

Have the most plausible findings been proposed, based on the preponderance of evidence considered?

Would the same findings be consistently arrived at by different interpreters of the same evidence?

Can these findings be reliably used for strategic purposes e.g. policy formulation, programmimg and budgetary decisions?

VALIDITY

Internal validity: The extent to which the evaluation came up with findings that directly respond to the evaluation questions in the particular case.

- The validity of the question
- The validity of the data – source, type, quality, comprehensiveness
- The validity of the methodology – design, collection, analysis

External validity: Are the findings applicable beyond the particular case?

SAMPLING

<u>Statistical sampling:</u> *Quantitative findings are representative of the tendencies in a population*

Should a program of activities directed at improving developmental outcomes of very young children be implemented nationally?

<u>Qualitative sampling:</u> *Representivity is focused on ensuring that all perspectives of consequence are considered*

Did a norm change program improve the attitudes and behaviour of adolescent males towards their female peers?

CASE STUDY: OVERVIEW

Assignment: Final evaluation of a comprehensive HIV care and treatment program

Key Constraint: 4 weeks of data collection

Purpose: Towards accountability

Design:
- Theory-driven rapid assessment
- Integrated analysis of mixed data
- Highly dependent on secondary data
- Primary data informs expert analysis, and supplements gaps if possible

CASE STUDY: KEY CAIR PROCESSES

The role of SME's is to identify, source and analyse the relevant data from an **informed** perspective

The role of the evaluation expert and team lead is to ensure that the team's analysis is **competent** (valid and reliable) and **astute** (integrating the influence of context)

Week 1 includes a team workshop in which each of the SME's must present answers to the following questions:

- What do we know about what works for achieving outcomes in your technical area in care and treatment programs? Present the evidence
- What do we know about how it works? Present the evidence
- How does the program we are evaluating align with what we already know?
- Based on the evidence, and what we know about the program we're evaluating, what can we expect to find?

By answering these questions we generate a program theory of change, an evidence map and a rubric for the evaluation
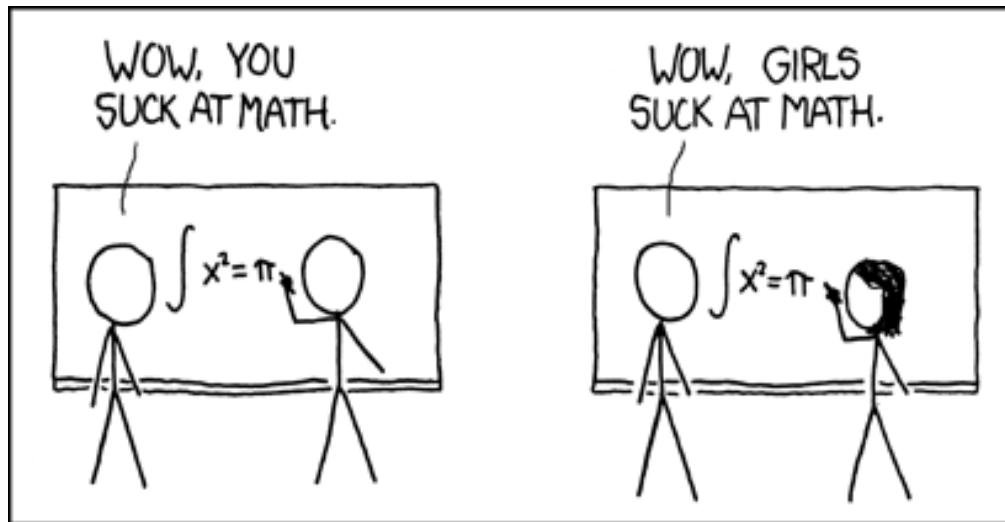
The team adopts a key **responsive** practice during the fieldwork – daily debrief facilitated by the evaluation lead

TECHNICAL EXPERTS ASSESSING RAPIDLY

" We're under a lot of time-pressure here, so we'll need to jump to conclusions. "

SMART PEOPLE JUMPING

THE RULES OF THUMB

The required burden of proof

The required burden of proof varies, depending primarily on whether the evaluation is being conducted for purposes of accountability, testing program efficacy, or other ends. Skillful evaluation takes its cue from the overarching purpose to makes its methodological choices.

The most plausible interpretation of the preponderance of evidence

Whatever the burden of proof required, an evaluation must present evidence that credibly supports its conclusions. There must be enough credible (reliable and valid) evidence to favor a conclusion above its alternatives.