EVALUATION, PART I: INTRODUCTION, AND MONITORING AND PROCESS EVALUATION
This lecture appears near the end of my course on Development Project Design and Evaluation.

The course is not a comprehensive guide to evaluation; rather, we aim to introduce the topic in sufficient detail that students who become development professionals can plan the outlines of evaluations and access the resources and expertise they will need to carry them out.

The lecture is very interactive, particularly from slides 22ff, where lots of student input is necessary to make it work well.
Monitoring vs Evaluation

Project Cycle Stage

1. Project Identification, Prep
2. Project Appraisal, Neg
3. Project Planning, Design
4. Implementation
5. Project Completion
6. Sustainability
7. Lessons Learned
Different types of monitoring

- Process evaluation (often focuses on implementation fidelity)
- Descriptive monitoring (rich qualitative data)
- Performance monitoring (ongoing measurements of project outputs)
Why is M&E necessary?

- Monitoring
  - Ensuring that resources are used efficiently
  - Assessing quality of project outputs
  - Assessing timeliness of project outputs
  - Identifying and correcting problems in implementation
  - Ensuring that project services and benefits are available to intended targets
  - Ensuring that project services and benefits are equitable across different groups
  - Assessing unintended or unforeseen consequences
  - Good monitoring makes evaluation easier
Key Monitoring Principles

The purpose of gathering data during project design and implementation is not simply to have the data.

Monitoring must be used as a tool for continuous project improvement. Data that isn’t used is worthless.
Key Monitoring Principles

Monitoring data can’t be collected haphazardly.

Data requirements and collection strategies must be identified and planned before the start of any project activities.
Key Monitoring Principles

It’s unethical to collect data that you don’t end up using.

Plan carefully, and answer these questions: Who is going to receive the data? What are they going to do with it? How will the data affect project implementation and management?
Key Monitoring Principles

Monitoring isn’t a “natural” part of project implementation. It doesn’t just happen.

Make sure it’s someone’s job to analyse data as it comes in.

Make sure that reports get written. Make sure that results get discussed.
Monitoring Data Sources

- Planned timetables, and actual implementation schedules
  - Data sources:
  - When delays are uncovered, you need to adjust your future plan

Table 4-2. Using a Two-Bar Gantt Chart to Plan and Monitor the Main Phases of a Housing Project

<table>
<thead>
<tr>
<th>Phase</th>
<th>1986 1 2 3 4</th>
<th>1987 1 2 3 4</th>
<th>1988 1 2 3 4</th>
<th>1989 1 2 3 4</th>
<th>1990 1 2 3 4</th>
<th>Revised Estimate (weeks)</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approval of contract</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+8</td>
<td>Completed</td>
</tr>
<tr>
<td>Earth moving</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+16</td>
<td>Completed</td>
</tr>
<tr>
<td>Installation of water</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+18</td>
<td>82% of domestic water connections completed</td>
</tr>
<tr>
<td>Installation of sewage system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+15</td>
<td>72% of sewage pipes laid</td>
</tr>
<tr>
<td>Installation of drainage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+13</td>
<td>44% of drains laid</td>
</tr>
<tr>
<td>Road construction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+6</td>
<td>Roads 65% complete but steep terrain will slow completion</td>
</tr>
<tr>
<td>Selection of participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+6</td>
<td>6,000 to 12,000 participants already selected</td>
</tr>
<tr>
<td>Construction of core housing units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+12</td>
<td>20% of core units constructed</td>
</tr>
<tr>
<td>Group house construction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+18</td>
<td>Lag due to delays in earlier stages</td>
</tr>
<tr>
<td>Construction loan approval</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+12</td>
<td></td>
</tr>
<tr>
<td>Plot occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+17</td>
<td></td>
</tr>
<tr>
<td>Community facilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+18</td>
<td></td>
</tr>
</tbody>
</table>

Note: Column headings indicate year and quarter.
Monitoring Data Sources

- Financial reports → do expenditures line up with expectations?
- Staff reports → Weekly reports, reports after major events, fieldnotes, etc.
- Participant lists → are people participating in the program as expected? Quantity? Is the target population showing up? Are other types of people asking for the services?
- Individual case management files
- Press coverage
Why is M&E necessary?

- Evaluation
  - Verify that project actually works
    - Did it do what it said it would do? (Think in terms of project budget, activities, outputs, objectives, goals).
  - Assess all the effects of the project
    - Positive and negative
    - Foreseen and unforeseen
  - Identify opportunities for scale-up, dissemination of the project
    - Identify challenges to scale-up
Key Evaluation Principle

Beyond satisfying the donor, evaluation is fundamentally about improving your project specifically and development practice generally.

Good evaluation also considers all the stakeholders in the process, and involves them in both the definition of the scope and process of evaluation.
The Challenges of M&E

- Good evaluation is expensive and time consuming.
- Projects are complex beasts, and it can be difficult to obtain, process, and integrate data from multiple sources.
- Projects take place in the real world. Integrating a transactional-ecological perspective into evaluation is difficult.
- It can be difficult to gauge the effects of a project over time. Knowing when to evaluate is difficult.
  - Getting a development agency to do an evaluation with multiple stages of measurement over a long time frame is almost impossible.
  - This can be a significant role for universities.
The Challenges of M&E

- Good evaluation requires both qualitative and quantitative data
  - There are increasing numbers of people trained in mixed-methods data collection
- Quant evaluation can involve advanced statistics
  - Contract out data collection, analysis
- Simpler evaluations have a difficult time defining mechanisms of causation
Love talks about the “transparent box” model of implementation evaluation, where the external forces and context which affect a project are made explicit. However, he doesn’t really talk about a process to evaluate whether our operational description of and assumptions about context are correct. How can we do that?
Group Work

Scenario: You are Associate Dean of Graduate Education here at Peabody (this is in the future, and Craig Anne has retired). You are tasked with putting together a system to monitor the process of graduate education at the college?

- What data would you need?
- From whom would you collect them?
- When would you collect it?
- How would you use it?
Individual Work

What data do you need to monitor the implementation of your project (the one you’re writing the proposal for)?

- How are you going to collect it?
- How are you going to analyse it?
- How do you see the monitoring process affecting your project?
- Use the chart on p110 of the book to think about your data points
Scenario: You are Associate Dean of Graduate Education here at Peabody (this is in the future, and Craig Anne has retired). You are tasked with putting together a system to monitor the process of graduate education at the college?

- What data would you need?
- From whom would you collect them?
- When would you collect it?
- How would you use it?
Discussion

- What’s the difference between monitoring and evaluation?
Developing an Evaluation

Scenario: Nonprofit JTB Water Services is fighting water-borne diseases in West Africa. After a two year education project in a village, it’s time for an end-of-project evaluation. JTB Water Services puts together a report with photo montage of the filters its distributed, the description of the satisfaction questionnaires it distributes after the end of its workshops, a highlights two case studies of its work. One case study is a man who says that JTB educated him not to wade in rivers or ponds when he was infected with guinea worm. Another case study was of a young mother who received information about the necessity of boiling water to protect her small children from disease.
What’s wrong?

- What could JTB Services do to improve the evaluation?
What’s wrong?

- JTB didn’t actually measure any project impacts.
  - Satisfaction reports aren’t project impacts, they just rate people’s satisfaction with services received. That’s important, but it’s not an impact per se.
    - There could be a social desirability bias, where people will tell the evaluator what they think he or she wants to hear.
  - The case studies give details only on the interventions of JTB, not the impact of those interventions.
    - So what if a man now knows not to wade in water when he has guinea worm? Does he actually follow that advice? What about other people who might have guinea worm?
JTB goes back to the drawing board. They decide that they’re going to measure four things:

- Prevalence of guinea worm infection
- Behaviors of people infected with guinea worm
- Prevalence of diarrhea infection among children <5y.o.
- Number of diarrhea-related deaths

JTB decides to collect data by putting together a survey that will be distributed to all the people that have participated in their programs.
Results

- Based on the surveys, JTB finds the following:
  - There have been 25 cases of guinea worm infection identified during the past six months in the 250 people who went through their guinea worm training program.
  - Of the 25 people infected with guinea worm, 20 say that they have been careful not to expose the worm to public water sources.
  - There have been 150 cases of diarrhea in the 300 children whose mothers have participated.
  - 10 children in that group have died as a result of diarrhea-related complications in the past six months.
Improving Evaluation

- What did JTB do better?
- What could be improved further?
Improving Evaluation

- **Good**: There are actual results to measure

- **Bad**: The rates of infection don’t actually tell us anything about the efficacy of the program
  - How do infection rates in the intervention group compare with people who haven’t participated in the intervention?
  - Guinea worm infection is a community problem and needs to be measured in the community
Improving Evaluation: Control Groups

- After reviewing the results, JTB realizes it needs to compare its data with data from other sources to know if its program is effective
  - One idea is to go to the local hospital and get data about infection rates etc from it
    - Hospital may not collect data systematically, may use different questions, and can’t know of cases that do not access its services
Another idea is to distribute the survey to other people in the community who have not participated in JTB programs and compare the results from the two groups.

This is a perfectly acceptable comparison group.

There is a danger that people who participated in JTB programs will tell their friends in the community about it who might change their behavior as a result.
Another idea: distribute the survey to a group of people in another village

- Purposefully selected to be isolated from the intervention
- Purposefully selected to be demographically similar to the intervention village
  - Geography
  - Distribution of ethnic groups
  - Similar economic profile
Improving Evaluation: Pre-Post Tests

- After reviewing the data from Village B and comparing it to Village A, JTB realizes that there are still some problems.
  - First, the composition of Villages A and B are different. Village A has a lot of Muslims, whereas Village B is mostly Christian. JTB doesn’t know if the differences they see in the survey results between the villages are due to its intervention in Village A or because of innate differences between the two.
  - Second, some of their informants from Village A say that they haven’t actually changed their behavior as a result of the intervention. Most of the content JTB offered as part of its educational program was already well known.
Pre-Post Tests

- One possibility to overcome this problem is to ask people in both villages how much their behavior has changed since the intervention.
  - Problems: Self-reported behavior is notoriously unreliable, and there is social desirability bias.
Pre-Post Tests

- JTB now realizes that it should have collected some baseline data in both villages prior to beginning the intervention.
- Post-intervention data would then be compared to the baseline data.
Quasi-Experimental Design: The Gold Standard

QEDs require at least one comparison group, at least one intervention group as well as pre- and post-intervention measurements of the outcome variables.

\[ \text{Project Impact} = I_{O_2} - I_{O_1} - (C_{O_2} - C_{O_1}) \]

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Intervention</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>(O_1)</td>
<td>Yes</td>
<td>(O_2)</td>
</tr>
<tr>
<td>Control</td>
<td>(O_1)</td>
<td>No</td>
<td>(O_2)</td>
</tr>
</tbody>
</table>
A Caution

- JTB needs to collect a wide range of demographic characteristics to check for differential outcomes across demographic groups.
  - In the aggregate a project may show no effect, but parsing results out by demographics may show systematically negative results for men and systematically positive results for women which average out to zero.
A Caution, Part II

- Relying on measurement of change (inherent in the pre-post method) privileges the unit of change

- Consider this example of results from a reading program

<table>
<thead>
<tr>
<th></th>
<th>Pre-Test Score</th>
<th>Intervention</th>
<th>Post-Test Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention--HS</td>
<td>50</td>
<td>Yes</td>
<td>65</td>
</tr>
<tr>
<td>Intervention—ES</td>
<td>20</td>
<td>Yes</td>
<td>30</td>
</tr>
<tr>
<td>Control—HS</td>
<td>50</td>
<td>No</td>
<td>53</td>
</tr>
<tr>
<td>Control—ES</td>
<td>20</td>
<td>No</td>
<td>22</td>
</tr>
</tbody>
</table>
Henry’s Comparison Groups

- Naïve → naturally occurring groups
- Value-added → include pre-test
- Regression-adjusted covariate → include demographic characteristics
- Value added + covariates → combination of 2 & 3
- Interrupted time series → pre-post w/ comparison group
- Fixed-effects design for longitudinal → no comparison group, measures during treatment and not
- Matching designs
- Regression discontinuity
Experimental Design:
The Platinum Standard (?

- QEDs are often criticized for relying on non-randomly selected intervention and control groups
  - Intervention groups may be people who have self-selected to be part of the intervention because of their inherent interests, etc
  - Control groups may be samples of convenience
- The federal government is increasingly mandating on true experimental designs
Experimental Design

- Experimental designs differ from quasi-experimental designs in that people are randomly assigned to control and intervention groups.
- Lists of people are compiled who are interested in participating, and people are randomly assigned to control and intervention groups.
  - Everyone on the list has an equal probability of ending up in either group.
Randomized Controlled Designs Issues

- Bias in allocation or follow-up
- Contamination
- Cross-over
- Attrition

Some alternatives

- Waiting list / stepped wedge design
Some Complications

- Sample size & statistical power
- Analytical expertise
Evidence-Based Practices

- Expert Panel Review of Research Evidence
- Meta-Analytic Studies
- Clinical Trial Replications With Different Populations
- Literature Reviews Analyzing Studies
- Single Study/Controlled Clinical Trial
  - Multiple Quasi-Experimental Studies
  - Large Scale Multi-Site, Single Group Design
- Quasi-Experimental
- Single Group Pre/Post
- Pilot Studies
  - Observational Experience
  - Established Clinical Practice
- Case Studies
  - Descriptive Writing
Discussion

- Do you see any ethical concerns in using experimental designs when implementing and evaluating development projects?
Case Studies

- A bounded unit of analysis:
  - Individuals
  - Programs
  - Communities

- One versus multiple cases
  - Selection questions: random versus purposive
Case Studies

- Importance of designing the process well at the beginning

- Data collection
  - Interviews
  - Document reviews
  - Direct observation
  - Program data
Use of stories

- Similar to case studies

- Used to:
  - Illustrate other data
  - Augment quant data
  - Identify patterns and trends
  - Offer insight on rare experiences
Use of stories

- Questions...
  - About confidentiality
  - About ‘truth’
  - About presenting stories to others
Identifying the right method

- What’s your in-house expertise?
- What can you afford to contract out?
- What kind of data can you reliably obtain?
- What can your consultants do well?
- What does your funder expect?
- What kind of information can your organization use effectively for program improvement?
- What kind of information is useful in fundraising?
Planning for your own projects

- If you had unlimited time and money, how would you evaluate the project for which you’re writing the proposal?

- What compromises would you be willing to make to the plan to make it faster and cheaper while still producing valuable findings?