Packing Your Bags with the Evaluation Toolkit

2007 Abstinence Education Evaluation Conference

Chris Trenholm
Mathematica Policy Research
Our Agenda

- Understanding Our Road Map
- Picking Our Vehicle
- Choosing Our Destination
- Getting Out of Our Driveway
- Driving Safely on Exit Three (Impacts)
Understanding Our Road Map: Types of Evaluation

- **Formative Evaluation**
  - Informs program development & implementation

- **Process (Descriptive) Evaluation**
  - Informs service delivery

- **Outcomes Evaluation**
  - Informs performance (monitor vs. benchmarks)

- **Impact Evaluation**
  - Informs program effects (determine causality)
Understanding Our Road Map: Evaluator Roles

- **Technical Assistance Evaluator**
  - Provides advice and recommendations on program development and improvement
  - Quick turnaround
  - Can often serve in a consultancy role

- **Co-Evaluator**
  - Teams with program staff having its own expertise
  - Can provide a range of evaluation support

- **External Evaluator**
  - Provides independent evaluation expertise
  - Can provide a range of evaluation support
## Choosing our Vehicle: Matching Evaluation and Evaluator

<table>
<thead>
<tr>
<th></th>
<th>T/A Evaluator</th>
<th>Co-Evaluator</th>
<th>External Evaluator</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Formative</strong></td>
<td>Good</td>
<td>Good</td>
<td>OK</td>
</tr>
<tr>
<td><strong>Process/Outcomes</strong></td>
<td>OK</td>
<td>Best</td>
<td>Good</td>
</tr>
<tr>
<td><strong>Impacts</strong></td>
<td>Poor</td>
<td>OK?</td>
<td>Best</td>
</tr>
</tbody>
</table>
Choosing Our Destination

Exit One:
Formative and Process

Exit Two:
Outcomes

Exit Three:
Impacts
Getting Out Of Our Driveway: Logic Model

Program Context → Program Activities

Program Outputs

Shorter-Term Outcomes → Long-Term Outcomes

Formative & Process Evaluation

Outcomes & Impact Evaluation
Antecedents of Teen Sexual Activity

Demographic Characteristics

Backgrounds & Experiences

Contextual Factors
  • Community
  • School

Services Available

Abstinence Education Program

Other Health, Family-life & Sex Ed Services

Services Received

Classes or Programs Addressing Key Topics

Programs or Meetings for Parents

Key Outcomes

Intermediate Outcomes

Views on Abstinence, Teen Sex, and Marriage

Peer Influences and Relations

Perceived Consequences of Teen and Nonmarital Sex

Behavioral Outcomes

Sexual Abstinence

Sexual Activity

Expectation of Future Behavior

Drug and Alcohol Use
Driving Safely on Exit Three (Impacts)

Key Concepts for Conducting Good Impact Evaluation
A Program Impact Is Estimated as the Difference Between Two Groups

- Mean outcomes of individuals participating (or having the opportunity to participate) in a program
  minus
- Mean outcomes of those same individuals had they not participated (or had the opportunity to participate) in that program
Q: How do I observe the mean outcomes of those same individuals had they not participated in that program (known as the “counterfactual”)?

A: You can’t. You can only estimate them. And that is often hard to do accurately.
How Do I Estimate the Counterfactual?
Experiments Are Ideal

- Defining feature: random assignment
  - assignment to groups based on random event
  - can assign either individuals or some pool of individuals

- Leads to two distinct groups of individuals
  - treatment group: selected for the program
  - control group: excluded from program (remain in the existing policy environment)
What Is So Great About Random Assignment?

- **The Control Group**: provides a reliable method to measure the counterfactual; that is, the outcomes that would have taken place without the program.

  Leaves only two reasons program group and control group outcomes differ:

  1. Program had an impact
  2. Random chance
How Do We Get Rid of Random Chance?

- Increase the study sample -- make the program and control groups bigger
- Whenever possible, conduct the random assignment among individuals not pools of individuals (e.g. classes, schools)
So Why Don’t We Experiment All The Time?

Conditions must be right!

– Must have “excess demand” for the program

– Cannot alter program delivery in important ways

– Need to address specific concerns as they arise
So What Do We Do When We Cannot Experiment?

- Conduct a “quasi-experiment” – i.e., obtain a comparison group that proxies really well for an experimental control group

- Many (many!) options but their credibility can often be questionable
Convenient Option: Pre-Post Methods

- Comparison group formed by outcomes of individuals in program group before they participate

- Is this a good proxy for a control group?
  - Rarely
    - followup must be short
    - testing effects can be problematic
    - attrition must be modest
Better Option: External Comparison Group

- Comparison group formed by outcomes of individuals with similar characteristics and experiences of individuals in program group

- Is this a good proxy for a control group?  
  - Sometimes  
  - Credibility improves by:  
    - using matching and/or regression  
    - measuring baseline outcomes for both groups
Common External Comparison Groups

- **Youth in same school who are eligible but do not enroll in a program**
  - They did not enroll for a reason (e.g. motivation)
  - Can easily cause impacts to be overstated

- **Youth in other schools who would be eligible otherwise**
  - They are subject to different environments
  - Can again be hard to sort out who would enroll

- **Youth in grade ahead who would have participated otherwise**
  - Can work well when program is new to school
  - Cohort differences are a main concern
What if Impact Estimates Lack Credibility?

- Be very clear about your concerns
- or -
- Return to an earlier exit
  - Avoid the analysis altogether
  - Conduct the analysis but treat results as process/outcomes findings
Ohhhh!!!!…
And Don’t Forget to Pack:

- Active Parental Consent
- Assure Data Confidentiality
- Established Survey Instruments