Introduction to Process Evaluation and Fidelity Assessment

Roger Przybylski
Research Director
Justice research and Statistics Association
rprzybylski@jrsa.org
What We Will Cover

• Nature of Evaluation
• Nature of Process Evaluation
  • What is its subject matter?
  • Why is it important?
• Process Evaluation Methods
  • Fidelity Measurement
• Q&A
Nature of Evaluation

• Old practice, new discipline
• Evaluators typically have been trained in other fields
  • Considerable debate about best ways to approach and practice evaluation
  • Competing paradigms
Nature of Evaluation

• Transdiscipline with many applications
  • Program, product, personnel, policy, proposal and performance are the big 6
  • Subject matter is merit, worth
    • Merit: intrinsic quality
    • Worth: extrinsic quality; quality in context

• Evaluation has a unique core logic

Facts + Values = Evaluative Conclusion

4
Values

• Which dimensions of merit/worth will be used
• Basis for determining good/poor performance on each dimension
• Basis for an evaluative conclusion
• Who is at the table; who decides
Nature of Process Evaluation
Focal Points of Evaluation

• Program conceptualization and design
• Program implementation/delivery
  • *Process evaluation*
• Program effects/outcome
  • *Outcome/impact evaluation*

Hierarchy of Questions

- Is the program economically viable?
- What outcomes were achieved?
- How was the program implemented/delivered?
- How does the program design address those needs?
- What are the needs of impactees?
Process Evaluation

• Documents how the program was implemented/delivered
• Discovers contextual factors
• Determines extent to which program was implemented/delivered according to plan
  • Program plan/model is normative reference
Why Process Evaluation is Important

• Essential for understanding/properly interpreting outcomes
• Provides blueprint for replication
• Critical for identifying implementation/delivery problems before they become intractable
Basic Evaluation Steps

• Engage stakeholders
  • Identify the evaluation’s purpose and audience

• Develop conceptual model of the program

• Focus the evaluation design
  • Develop evaluation questions
  • Choose design/methods

• Gather credible evidence
  • Collect and analyze data

• Report findings
Real World Outcomes/Benefits are Shaped by Intervention Effectiveness and Implementation Effectiveness

“The ideas embodied in innovative social programs are not self-executing.” Instead, what is needed is an “implementation perspective on innovation—an approach that views post-adoption events as crucial and focuses on the actions of those who convert it into practice as the key to success or failure.” (Petersilia, 1990)
Key Terms

• Program design: a plan stating expectations about how the program will work

• Implementation fidelity: extent to which a program is implemented/delivered according to design

• Outcome: changes or benefits produced by the program in the target area or population
Group Exercise
Program Option 1
28% Negative UAs

Program Option 2
82% Negative UAs
Juvenile Intervention

- Business as Usual: 28% Recidivism Rate
- Program Option 1: 34% Recidivism Rate
- Program Option 2: 18% Recidivism Rate
Washington State Institute for Public Policy Evaluation of Functional Family Therapy in Washington State

Source: Adapted from Barnoski (2004)
Implementation Matters

- Identifying and adopting what works is not enough to achieve successful outcomes
  - Evidence-based initiatives still have to be delivered with fidelity/integrity in diverse and complex real-world settings

One of the strongest messages coming from the research is that fidelity—the quality with which the treatment is delivered—is crucial to successful outcomes. Lipsey et al. (2010)
Implementation Gap in Corrections

• Studies indicate problems in the administration of risk assessment tools and the development of correctional case management plans are common
  • In a survey of probation practitioners, Miller & Maloney (2013) found that only about one-half of the respondents required to use structured risk assessments completed the tool carefully, making decisions congruent with tool recommendations

Several studies have found discordance between an offender’s needs and the treatment targets specified in case management plans, suggesting that lack of adherence to the need principle is a common problem

(Haqanee et al., 2015; Viglione et al., 2015; Bonta et al., 2008)
Implementation Gap: Adherence to EB Principles Matters

• Using an RCT design, Marques et al. (2005) evaluated the effectiveness of treatment for sexual offenders and found that treatment did not reduce recidivism
  • The researchers also reported that the treatment program did not adhere to the principles of RNR
• In a subsequent meta-analysis, Hanson et al. (2009) found that sex offender treatment did reduce recidivism, and that adherence to the RNR principles moderated treatment effectiveness
High-Quality Implementation is Difficult to Achieve

• Ordinary circumstances present serious obstacles to high-quality implementation
  • What appears to be simple and straightforward in the implementation process often turns out to be more complex than anticipated
  • We often underestimate the number of steps involved, the number of separate decisions that have to be made, or the number of participants whose preferences have to be taken into account

(Pressman & Wildavsky, 1973)
Purpose of Evaluation

**Formative**
- Feedback for program improvement
- Program management tool
- Flexible, often interactive plan
- Periodic reports, responsive to client requests

**Summative**
- Final report card on effectiveness
- Typically for benefit of external audience
- Fixed plan
- Findings typically reported in a final, all-inclusive report
Developmental Evaluation

• Supports innovation
• Logic/conceptual model unknown or under development
• Informs development from design through expected outcomes
• Highly collaborative

• Encourage you to watch:
  https://www.youtube.com/watch?v=jB6cdoR16v0
Program Conceptualization and Design

• Key questions
  • What is the need or demand for the program? What problem is the program intended to address?
  • Who are the intended beneficiaries or targets of the program?
  • What is the scope of the initiative in terms of its geographic boundaries or number of intended beneficiaries?
  • What are the underlying assumptions or theories that define the relationships or chain of events that are expected to lead to intended outcomes?
Logic Model

• Graphic representation of the way a program is intended to work
• Depicts relationship between needs, program activities and expected results
• Specifies the program’s theory of action
• Provides common understanding of program
Logic Model
A Simple Example

Examined to ensure that components are logically linked

Problem or Need
- Headache
- Grouchy
- Can’t Work

Activities
- Take Aspirin
- Follow Dosage Rules

Intermediate Outcomes
- Headache Gone

Long-Term Outcomes
- Less Grouchy
- More Productive
Logic Model
A Program Example

Examined to ensure that components are logically linked

Problem
- High Recidivism
- Poor Job Skills

Activities
- Vocational Training
- Job Referral

Intermediate Outcomes
- Improved Job Skills
- Gainful Employment

Long-Term Outcomes
- Lower Recidivism

Helps identify what needs to be measured

Mechanism
- Learn Skills
- Apply/Interview Successfully

27
## Logic Model: Key Elements and Definitions

### Key Terms

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Inputs</th>
<th>Activities</th>
<th>Outputs</th>
<th>Outcomes (Intermediate)</th>
<th>Impact (Long-term)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem, need, threat, or gap in service</td>
<td>Resources available to conduct program activities</td>
<td>Actions, services, or interventions put in place to achieve program goals</td>
<td>Counts of the internal activities of a program</td>
<td>Changes in the lives or circumstances of persons or an area served or targeted by the program</td>
<td>Long-term changes in individual or group behavior or community conditions</td>
</tr>
<tr>
<td></td>
<td>May refer to funds, people, facilities, partnerships, and plans</td>
<td></td>
<td>Products or units of service delivered</td>
<td>Reflects events or conditions external to the program and of direct importance to the public</td>
<td>Durability of change over time</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Side effects</td>
</tr>
</tbody>
</table>
## Logic Model: Relationship to Evaluation and Performance Measurement

<table>
<thead>
<tr>
<th>PROCESS EVALUATION</th>
<th>OUTCOME/IMPACT EVALUATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conditions</strong></td>
<td><strong>Inputs</strong></td>
</tr>
<tr>
<td>Problem, need, threat, or gap in service</td>
<td>Resources available to conduct program activities</td>
</tr>
<tr>
<td></td>
<td>May refer to funds, people, facilities, partnerships, and plans</td>
</tr>
</tbody>
</table>

### Key Performance Indicators in a Sound Performance Measurement System
Process Evaluation Methods
A process evaluation might attempt to answer the following types of questions:

- How is the program organized and staffed?
- Who participates in the program?
- How is the screening/selection process for program participation carried out?
- Do program participants differ from the target population specified in the program’s design? If so, how?
- What services or interventions are provided?
- What is the intensity and duration of each service or intervention?
- How are these services or interventions delivered? What is the quality of delivery? Are they delivered in a consistent manner?
- Do all participants complete the program? If not, who completes the program and how do they differ from who drops out?
- Was the program implemented/delivered according to plan?
Answering Process Questions

• Both qualitative and quantitative methods are typically used
• Qualitative data obtained through *interviews* with program staff, stakeholders and sometimes clients, as well as through *on-site observation* of program activities and a *review of program documents/records*
• Typically supplemented with quantitative data obtained from program records
  • Inputs, outputs, coverage, attrition
• Combine information from different sources to produce an accurate description of the way the program is operating in real-life
Evaluation Design/Methods

Overall plan and specific techniques for answering the evaluation questions — Many options, all have strengths and weaknesses

- Quantitative
  - Hard data; facilitates causal analysis
  - Presents narrow picture, reduces complex world to numbers

- Qualitative
  - Natural/holistic view, deeper understanding
  - Open to interpretation/bias, can be difficult to analyze
Quantitative Inquiry

- Experiments, quasi-experiments, hypothesis testing, significance tests, descriptive and inferential statistics
- Sampling is random, representative
- Common data collection methods include surveys, questionnaires, tests, accessing official statistics and program records
- Data produced: numbers, statistics
- Analysis: descriptive statistics, relationships between variables, causal analysis, inferential statistics, significance testing

Qualitative Inquiry

- Ethnography, participant observation, case studies
- Sampling is purposive: i.e., critical case, variation, snowball
- Common data collection methods: interviewing, direct observation, document review
- Data produced: quotations, descriptions, excerpts
- Analysis is descriptive, interpretive; often looks for themes, patterns, insights, explanations
Scientific Rigor

Criteria for Judging Rigor

Quantitative

- Construct Validity
- Internal Validity
- External Validity
- Reliability

Sound operational measures for the concepts being studied.

Ability to attribute effects to the program rather than something else:
Confidence that the program caused the effect.

Ability to generalize findings:
Confidence that the program will work in another setting.

Several classes: i.e., inter-rater, test-retest. Replicability or repeatability.
Different observers should score the same phenomena the same way.
Scientific Rigor

Criteria for Judging Rigor

Quantitative
- Construct Validity
- Internal Validity
- External Validity
- Reliability

Qualitative
- Credibility
- Transferability
- Dependability
- Confirmability

Prolonged Engagement, Triangulation
Thick Description
Audit
Fidelity Measurement

• Several models/frameworks
  • Critical elements/differentiation
  • Structure (framework for service delivery) and process (way in which services are delivered) (Mowbray et al., 2002)

• Common measurement/assessment themes
  • Context
  • Compliance: adherence, dosage
  • Competency
  • Responsiveness
Structure/Process Framework

**Structure**
- What resources are allocated to the program?
- Who are the key actors and partners in the program and what are their roles?
- Are key program elements in place?
- Who actually participates in the program?
  - Selection process, case flow, coverage and attrition
- What activities (services/interventions), both planned and unplanned, did the program undertake?
  - Adherence
  - Dosage
  - Quality, consistency, responsiveness
- What outputs were produced?

**Process**
Answering Process Questions

• **Qualitative methods play key role**
• Data collection protocols/training are key
• Purposive sampling
  • Prolonged engagement
• Quantitative component
  • Inputs, coverage, attrition
Fidelity Measurement: Inputs and Outputs

• Inputs: Resources used by a program, including staff, equipment and financial resources. Input measurement can be important not only for assessing a program’s implementation fidelity, but also for establishing the foundation for economic evaluation.

• Outputs: Direct products and services delivered by a program over a given period of time.
  • Reflect the internal activities of a program and are usually expressed as counts of workload or service delivered. Outputs reflect immediate results, which are internal to the program. Outputs should lead to desired program outcomes.
Fidelity Measurement: Program Coverage

• Coverage refers to the extent to which the target population participated in the program according to the program design.
• Under-coverage
• Over-coverage
• Bias
• Attrition
Fidelity Measurement: Key Measures/Focal Points

• Adherence: whether a program service or intervention is being delivered as it was designed or written

• Exposure or dosage: amount of an intervention received by participants; in other words, whether the frequency and duration of the intervention is as full as prescribed by its designers

• Quality of delivery; the manner in which staff delivers a program (perhaps the most ambiguous/difficult measure)

• Responsiveness: How well participants (and staff) respond to, or are engaged by, an intervention
Staff Responsiveness

• Staff training...when layered atop individual values and political environments that are philosophically contrary to the underpinnings of clearly articulated evidence-based practices, is ineffective.

• The values and belief systems of individual correctional practitioners and organizational cultures must be concerns of the first order.... The importance of the relationship between committed and competent leaders and successful program implementation cannot be overstated.

(Paparozzi & Guy, 2013)
Moderators of Fidelity

- Intervention complexity
- Facilitation strategies

- Also consider Collaboration
Process Evaluation Challenges

• Adaptation issues
  • Common, but remains controversial
• What constitutes high-fidelity and full vs. partial implementation?
• Metrics/tools for measuring quality of delivery/competency
• Empirical guidance regarding dosage is limited
  • Definition of dosage
  • Sequence of dosage
Data Collection and Management

• Identify all data needs when planning the evaluation
• Know how you will get the data
  • Source, collection method, timing
  • Sampling issues
• Establish quality control procedures
• Privacy and security issues
Thanks for Your Attention

Questions?

Comments?