Briefing Overview

- Introduce and explain the key concepts in outcome evaluation research.
- Goal - Help practitioners
  - Understand key evaluation terms and designs; and
  - Recognize how to identify a well-written evaluation report.
- Provides basic information needed to identify high quality evaluation reports
What is Evaluation?

- Evaluation uses a systematic and objective process to determine the impact of a policy or program.
- Evaluation is a type of research.
- Addresses whether and to what extent the program is achieving goals and objectives and the impact of the intervention.
- Practitioners can use the information to implement program improvements.
Types of Evaluation

- Generally categorized as being either process or outcome.
- Process evaluations focus on the implementation of the program or project.
- Outcome evaluations (impact evaluations) focus on the effectiveness of a program or project.
Evaluation Quality

- Evaluations vary widely in quality.
- Reading an evaluation report, particularly one full of statistics and technical terminology can be overwhelming.
- Two key issues to help distinguish between good and poor evaluations:
  - The quality of the evaluation design; and
  - How well the evaluation is carried out.
Issue 1: Role of Evaluation Design

- The research design is the most important piece of information to use in assessing whether the program had the intended impact.
- Must establish a causal relationship between the program and any outcome shown.
  - The cause preceded the effect;
  - The cause and effect are related; and
  - The effect was not caused by an outside factor.
Accounting For External Factors

- Successful School Anti-Bullying Campaign
- Homework Club
- Tutoring Program

Outcome: Improved Academic Performance
Research Designs: Factors

- Evaluation design is the structure of evaluation
- Factors to consider:
  - Resources (time and money)
  - Appropriateness
- Common Approaches:
  - Experimental
  - Quasi-Experimental
  - Non-Experimental
Experimental Designs

- Random assignment of subjects to
  - Treatment (received program/policy)
  - Control (did not receive program/policy)
  - Referred to as “Randomized Control Trial” (RCT)

- Considered “Gold Standard”
  - Assumes difference between groups is one received the treatment and the other did not.
Comparing Treatment to Control

- Once randomly assigned, treatment provided, then outcomes for each group are compared.

- A common design is the pre-post design:
  - Same Data are collected for each group before (pre-test) and then again after (post-test) the intervention.
  - Determine if there are differences in the degree of change from pre-test to post-test for each group.
Strength of RCT

- With an RCT design, little question about what caused changes observed in treatment group because approach controls for external factors.
- External factors are factors other than the program that may have caused the outcomes.
- An experimental design addresses these types of concerns, since both the program participants and non-participants would be exposed to the external factors (e.g., the school campaign).
Understanding the Report: Tip #1

- Design must be implemented well. Requires:
  - Assignment is actually random;
  - Control group does not receive the program; and
  - the evaluation examines outcomes for everyone, even those in the treatment group who don’t finish the intervention

- A good evaluation report on an experimental design should discuss these issues.
Quasi-Experimental Designs

- Difference: quasi-experimental does not employ a randomly assigned control group.
- Instead, uses a comparison group of those considered similar to the treatment group.
- To the extent possible, Evaluator may attempt to ensure both groups are comparable by matching on a variety of factors (e.g., age, gender, prior history).
Types of Quasi-Experimental Designs

- Pre-Post with treatment and comparison group most common
- Time Series – often used to assess effect of a new policy or legislation in a jurisdiction.
  - Evaluator conducts series of observations prior to the policy and then again after the policy.
Example Presentation Time Series

Number of Unexcused Absences by Quarter

Date of New Truancy Policy
Understanding the Report: Tip #2

- Depending on the program or policy under study, it may be impossible or undesirable to use an RCT design, or to identify an appropriate comparison or control group.
  - Particularly with programs/policies that affect an entire community or jurisdiction. Keep this in mind when considering the strength of the design.
- Always look for an explanation of why the evaluator chose a particular research design.

NJJEC
A Project of the Justice Research and Statistics Association
Non-Experimental Designs

- Provides the least amount of guidance
- Considered the weakest of all designs
- Raises the greatest amount of questions
- Evaluator collects data on the treatment group or those affected by the program/policy at one point only: after the program/policy occurred.
- Cannot assess if change occurred
Issue 2: Evaluation Execution

- Evaluator must ensure basic design is not undermined and other elements such as:
  - Data collection;
  - Analysis; and
  - Writing up the results
- Must be carried out sufficiently well so that one can trust the results
Sample Size

- Number of subjects (individuals, cases) in the study is large enough to support analysis and raise little doubt that the results adequately represent the population (the entire group from which the sample was drawn).

- Sample size is a key issue when one does not have the resources to permit data collection from all subjects in the population.
## Sample Size Guide

<table>
<thead>
<tr>
<th>Number in Population</th>
<th>Minimum Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>50</td>
<td>44</td>
</tr>
<tr>
<td>70</td>
<td>59</td>
</tr>
<tr>
<td>100</td>
<td>80</td>
</tr>
<tr>
<td>150</td>
<td>108</td>
</tr>
<tr>
<td>200</td>
<td>127</td>
</tr>
<tr>
<td>500</td>
<td>217</td>
</tr>
<tr>
<td>1000</td>
<td>278</td>
</tr>
<tr>
<td>1500</td>
<td>306</td>
</tr>
<tr>
<td>2000</td>
<td>322</td>
</tr>
</tbody>
</table>

Attrition

- Subjects dropping out of evaluation before data collection concluded
- Individuals don’t usually drop out randomly
  - Youth with lower grades more likely to skip school or more likely to drop out of program
- Result: Comparison and Treatment Groups no longer look similar
Understanding the Report: Tip #3

- Good evaluation studies track who drops out, when they drop out, and why they drop out.
- The number of subjects who drop out as well as descriptive information about the dropouts should be included in the evaluation report.
- Outcome information should be included on these subjects and should be a consideration in assessing the sufficiency of the sample size.
- Always look for a discussion of the sample size, attrition, and if outcomes include drop-outs.
Implementation

- Problems in implementation can result in the treatment group being substantially different than the comparison group. Such as:
  - Recruitment of subjects
  - Assignment to treatment and control not occurring as planned
- Subjects may then differ on key characteristics (e.g., age of youth, school attendance history)
Outcome Questions

- Main outcome question is often
  1. Did the treatment group (those that received the program/were impacted by the policy) have better outcomes than those who did not receive/not affected by the program/policy? or
  2. Did the treatment group improve/change since the intervention was delivered?
- Expectation that certain subgroups of subjects would have different outcomes.
Understanding the Report: Tip #4

- Specify the important characteristics of similarity between the treatment and comparison groups and whether the two groups are actually similar on these characteristics.

- The evaluation report should indicate whether or not analyses breaking out outcomes by subgroups were conducted and the rationale for doing so.
Outcome Measures

- Measures, or indicators, selected for assessing outcomes of the program/policy should:
  - Fit well with the program/policy objectives;
  - Accurately measure the concepts of interest.

- If delinquency is the outcome of interest, one could measure this in various ways, including:
  - Whether youth had ever been adjudicated; or
  - Number of times youth reported committing delinquent acts (e.g., underage drinking, theft)
Reliability and Validity

- Need to consider, when deciding how to measure delinquency, how good of a job the indicator will be in measuring delinquency – referred to as reliability and validity.
  - Reliability = consistency
  - Validity = Accuracy
- Report should explain why measures selected, process for selecting and time frame measured (e.g., within 6 months of program discharge)
Also consider the impact of the data selected to measure outcomes.

- Number of times arrested will produce a higher delinquency count than being sentenced to out of home placement;

- Precision is also important
  - Youth ever truant vs.
  - Number of times truant
Timing of Follow-Up

- When reviewing outcome reports, consider the timing used for follow-up.
- 3 months post-placement vs. 12 months post placement = different picture
- Follow-up period should be the same for all study participants
Understanding the Report: Tip #5

- Are measures suitable? Report should explain *why* the measures were selected, *the process* for selecting the measures, and the *time frame* measured.

- Do follow-up periods make sense? Evaluators should explain *why* they selected the particular follow-up time period; time should be *comparable* for all study participants.
More difficult are assessing whether the statistical tests used were appropriate and interpreted correctly (assumes quantitative study)

Statistical Significance = the size of the size of the observed differences between the treatment group and comparison group on the outcome was large enough that it was unlikely to have occurred by chance.

Without this, unable to know how much importance to place on the size of the observed differences.
Understanding the Report: Tip #6

- As with the selection of measures, the report should clearly indicate:
  - Why a particular statistical test was selected; and
  - How it was interpreted
Beyond Basics: Cost-Benefit

- Some evaluators assess economic implications of the program. Two approaches are:
  - Cost-effectiveness and
  - Cost-benefit analysis (CBA)
- Cost-effectiveness analysis examines the monetary costs and outcomes of a program.
- CBA assesses not only the costs and effects of the program, but also the benefits and whether the benefits outweigh the costs.
- See [http://cbkb.org](http://cbkb.org) for more CBA information
Beyond Basics: Meta-Analysis

- Meta-Analyses and Systematic Reviews synthesize the results of studies on a particular topic in order to produce a summary statement on the question being examined.
- Typically only include studies with strong research designs that were carried out well.
- These studies are important because they can be used to determine whether a program is evidence-based.
Meta-Analysis (Cont)

- As procedures specified, the approach should be able to be replicated by others.
- Terms are not necessarily interchangeable because of the methods used, meta-analyses and systematic reviews generally have the same purpose.
- Be aware that both meta-analyses and systematic reviews include a much more complex, thorough, and methodical review than is conducted by a literature review.
Putting It All Together

- Takes time to assess quality of an evaluation report.
- Valuable investment knowing whether a report is of good quality will help with issues such as contradictory statements (e.g., one evaluator states the program works, while another states the program does not work).
- Guide not comprehensive; far more can be done but this should help the practitioner make informed decisions.
Definitions of Key Evaluation Terms

- **Sample Size**: Number of subjects in the study.
- **Sufficient sample size**: Number of subjects is large enough that there is confidence that the results represent the population being studied.
- **Statistical significance**: Whether the program/policy is likely to have caused the desired result/change.
- **Effect size**: How much of a change the program/policy caused.
Resources

- The Research Methods Knowledge Base: http://www.socialresearchmethods.net/kb/
### Understanding Juvenile Justice Evaluation Reports: Issues To Consider

<table>
<thead>
<tr>
<th>What is the research design (experimental, quasi-experimental, non-experimental)?</th>
<th>Is the comparison/control group comparable to the treatment group (if applicable)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did random assignment go as planned (if applicable)?</td>
<td>Do analyses appropriately examine outcomes by subgroups (i.e., break out outcomes by subgroups)?</td>
</tr>
<tr>
<td>Did anyone in the control/ comparison group receive the intervention (if applicable)?</td>
<td>Are the measures suitable?</td>
</tr>
<tr>
<td>Did the evaluation examine outcomes for program dropouts?</td>
<td>Was there a comparable follow-up time for all evaluation subjects (if applicable)?</td>
</tr>
<tr>
<td>Is the sample size sufficient?</td>
<td>Did follow-up time make sense (if applicable)?</td>
</tr>
<tr>
<td>Does the report address attrition well?</td>
<td>Were statistical tests appropriate and interpreted correctly?</td>
</tr>
</tbody>
</table>
About NJJEC

- NJJEC is a project of the Justice Research and Statistics Association funded by the Office of Juvenile Justice and Delinquency Prevention (OJJDP). NJJEC’s purpose is to improve the evaluation capacity of states, localities, and tribes and facilitate the use of evidence-based programs and practices in juvenile justice.

Notice of Federal Funding and Federal Disclaimer

This presentation was prepared under Grant Number 2010-JF-FX-0063 from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. Points of view or opinions in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.