Rapid Learning: Methods for Testing and Evaluating Change in Social Programs

OPRE Methods Meeting
October 25, 2018
Introductions & Overview

• Why this work matters
• Getting started
• Methods
• Examples
• Final takeaways
Lesson #1: Move past the jargon
Lesson #1: Move past the jargon

Motivation: Transparency and Accountability
- Performance measurement
- Quality assurance
- Compliance reporting

Motivation: Practice change
- Nonprofit & public sector
- Health & public health
- Education
- Performance management
- Quality improvement
- Data driven decision-making

Data reported EXTERNALLY

Data used INTERNALLY

Blumenthal and Kilo 1998; Coburn and Turner, 2012
Lesson #1: Move past the jargon

- Start with program **objectives**
- Identify a **strategic** change to test
- Hypothesize potential improvements
- Determine the appropriate **measurement**
- Analyze the outcomes and make decisions
Lesson #2: Start with the question

What are you trying to answer?

METHOD ➔ QUESTION

QUESTION ➔ METHOD
Lesson #2: Start with the question

Identify the question

• What am I trying to understand?
• What outcomes do I want to change?
• How will I use the results? Who will use them?
• What are the organization’s priorities, and where does this issue fit?
• How confident do I need to be in the results?
• How hard is it to implement the innovation?
• How much will-building and engagement of program delivery staff do I need?
• What data are available?
• How soon do I need to know the results?
• What time is needed to observe an impact? Is there a near-term proxy for that impact?
Lesson #2.1: What am I trying to understand?

- How to apply established evidence
Lesson #2.1: What am I trying to understand?

- How to **apply** established evidence
- How to **adapt** established evidence to new context
Lesson #2.1: What am I trying to understand?

- How to **apply** established evidence
- How to **adapt** established evidence to new context
- How to **discover** new evidence
Lesson #2: Example

Highest priority considerations in 2006

- What am I trying to understand?
  Impact of professional development for preschool teachers

- What outcomes do I want to change?
  Classroom quality AND children's language and literacy

- How will I use the results? Who will use them?
  "to influence policy" "policy-makers"

- What are the organization's priorities, and where does this issue fit?
  To be a national leader improving the quality of public early education and children's opportunities

- How confident do I need to be in the results?
  Completely confident. Premium on experimental design.

- How hard is it to implement the innovation?
- What data are available?
- How soon do I need to know the results?
- What time is needed to observe an impact? Is there a near-term proxy for that impact?
Lesson #2: Example

Highest priority considerations in 2011

• What am I trying to understand?
  How to change the intervention (professional development for preschool teachers) to maintain or improve impacts on quality AND achieve impacts on child-level language and literacy outcomes

• What outcomes do I want to change?
  Classroom quality AND children’s language and literacy

• How will I use the results? Who will use them?
  To redesign the intervention; the implementation team and program recipients

What are the organization’s priorities, and where does this issue fit?
  To develop a more effective intervention. [To be a national leader improving the quality of public early education and children’s opportunities]

• How soon do I need to know the results? 1 to 2 years

• How hard is it to implement the innovation?
• How confident do I need to be in the results?
# Selecting Comparisons

<table>
<thead>
<tr>
<th>HISTORIC TRENDS</th>
<th>PERFORMANCE TARGETS</th>
<th>INDIVIDUAL CLIENT HISTORY</th>
<th>NATIONAL BENCHMARKS</th>
<th>PARTICIPANTS</th>
<th>NONPARTICIPANTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can we improve outcomes over time?</td>
<td>Can we better meet program targets?</td>
<td>Do participant outcomes improve?</td>
<td>Can we improve relative to the nation?</td>
<td>Do participants with the change have different outcomes?</td>
<td>Does the overall program improve outcomes?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Identify optimal comparisons**

- Can we improve outcomes over time?
- Can we better meet program targets?
- Do participant outcomes improve?
- Can we improve relative to the nation?
- Do participants with the change have different outcomes?
- Does the overall program improve outcomes?
## Selecting your Method

<table>
<thead>
<tr>
<th>Method</th>
<th>Level of Confidence</th>
<th>Methodological Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmarks/Targets</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Statistical Control Charts</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Comparative Interrupted Time Series</td>
<td>Moderate-High</td>
<td>Moderate-High</td>
</tr>
<tr>
<td>Matched Quasi-Experimental Design</td>
<td>Moderate-High</td>
<td>High</td>
</tr>
<tr>
<td>Prediction</td>
<td>Moderate-High</td>
<td>High</td>
</tr>
<tr>
<td>Randomized Control Trial</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Adaptive Randomized Control Trial</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

Timing varies for each method.
Measuring Improvement

- What data are available?
- What time is needed to observe an impact? Is there a near-term proxy for the impact?
Measuring Improvement

- What data are available?
- What time is needed to observe an impact? Is there a near-term proxy for the impact?

**Outcome:** Children’s early language and literacy skills

**Process:**
- Instructional Time
- Instructional Quality

**Near-term proxy:** ?

**Plan-Do-Study-Act**

**Model for Improvement**

**What are we trying to accomplish?**

Improving students’ language skills, sp. vocabulary

End-of-year Woodcock-Munoz

**How will we know that a change is an improvement?**

Daily measures:
- N children using new word with help;
- N children using new word without help

**What change can we make that will result in improvement?**

Introduce 1 new vocabulary word every day with rotating strategies for learning the new word

**Question to answer with PDSA Cycle #1:**

If we introduce 1 new vocabulary word each day, using rotating instructional strategies, will the number of children who can use new words without help increase with practice?

**Plan.** The teacher and aide will introduce a new vocabulary word every day, using different strategies each day. The teacher will ask children to use the word; the aide will track the number of children who use the word with help and without help.
Plan-Do-Study-Act

Do. The teacher & aide did introduce new vocabulary words each day, using rotating strategies.

Study. Daily data, weekly reflection.
• Over time, more children were using new words without help.
• Some words harder; more children needed help with them.

Act. Adapt: *re-introduce difficult words again later in the same week*, using a different instructional strategy.
Measuring Improvement

Near-term proxy – **Daily.**
Participants collect in PDSAs.

- Average N minutes per day dedicated to language activities

Near-term proxy – **3 times per year.**
Participants assess students.

- Average CLASS scores based on full-day video taping, 2014-2015

Process Measure – **Daily.** Participant report.
Comparison: self & peers over time.

- Process Measure – validated assessment.
Comparison to self & peers **3 times per year.**

- CLASS Scores
  - A. Socioemocional
  - O. Aula
  - A. Pedagógico
Measuring Improvement

Quasi-experimental analysis of internationally-validated language direct assessments. Comparison to non-participants.

Every 2-3 years.
Example: How might we?

- Individuals who have completed detox are at high-risk of relapse once back in their communities
- Post-detox programs can help reduce the number of relapses
- Program drop-out rate is high
### Comparing the Questions

<table>
<thead>
<tr>
<th>Question</th>
<th>Scenario #1</th>
<th>Scenario #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>What am I trying to understand?</td>
<td>Can the components of an effective framework…</td>
<td>…be streamlined…</td>
</tr>
<tr>
<td></td>
<td>…be adapted…</td>
<td>…to reduce dropout?</td>
</tr>
<tr>
<td>What outcomes do I want to change?</td>
<td>Substance abuse relapse</td>
<td></td>
</tr>
<tr>
<td>How will I use the results? Who will use them?</td>
<td>Program director, program staff</td>
<td>Program director</td>
</tr>
<tr>
<td>What are the organization’s priorities, and where does this issue fit?</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>What’s the risk of a false positive?</td>
<td>Low</td>
<td>Moderate</td>
</tr>
<tr>
<td>How hard is it to implement the innovation?</td>
<td>Low</td>
<td>Moderate</td>
</tr>
<tr>
<td>How much will-building and engagement of program delivery staff is needed?</td>
<td>Moderate</td>
<td>Limited</td>
</tr>
<tr>
<td>What data are available?</td>
<td>Participant history, risk predictions, daily attendance, program completion</td>
<td></td>
</tr>
<tr>
<td>How soon do I need results?</td>
<td>9 months</td>
<td></td>
</tr>
<tr>
<td>What time is needed to observe an impact? Is there a near-term proxy?</td>
<td>6 weeks. Yes.</td>
<td></td>
</tr>
</tbody>
</table>
Comparing the Timeline

- **Month 0**: Design a theory and measures
- **Month 1**: Recruit programs
- **Month 2**: Learning Session #1
- **Month 5**: Learning Session #2
- **Month 8**: Learning Session #3
- **Month 9**: Finalize lessons learned

**BSC**
- Design a theory and measures
- Recruit programs

**RCT**
- Design a theory and measures
- Recruit programs

**Run intervention**
- Program Administrators and Researchers review the data

**Program Administrators and Frontline Staff**
- Review the data
Contrasting the Approaches

<table>
<thead>
<tr>
<th></th>
<th>BSC</th>
<th>RCT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monitor:</strong></td>
<td>Daily attendance, retention to 5 visits, time between visits, program completion. Annotate adaptations</td>
<td>Daily attendance; fidelity</td>
</tr>
<tr>
<td><strong>Outcome:</strong></td>
<td>Daily return rate to the program, program completion rate</td>
<td>Program completion rate</td>
</tr>
<tr>
<td><strong>Question answered:</strong></td>
<td>Did the drop out rate shift significantly after we developed, refined and implemented a new process? By how much? Which adaptations were associated with the shift?</td>
<td>Is one combination more effective than the other? If so, by how much? What’s the cost-effectiveness of each combination?</td>
</tr>
</tbody>
</table>

**Pros**

**Cons**
Final Takeaways

- Let the context determine the method
  - Including decisionmaker needs, staff readiness, confidence required

- Factors that drive timeline are context-specific
  - Including time-to-impact, data availability, sample size

- Keep learning and improving
Questions?