



# How Do You Know that You are Making Progress? Evaluating Your Professional Development Program

Iris R. Weiss

# Evaluation is helpful in...

1. Continuous improvement of professional development programs
  - Deciding what mid-course corrections are needed
2. Making the case for professional development
  - Demonstrating the impact of professional development programs on teaching and learning

# Individual Reflection: History of the Future

- Think about a specific professional development program. Imagine that the program is wildly successful.
- Write “bullets” for the press release about the program’s success. (10 minutes)
- Discuss at your table. (15 minutes)

# Did your press release include:

- Information on the professional development design?
- Claims of impact?
- Evidence to back up the claims?

- What data you collect, from whom, and who does the analysis, interpretation, and reporting depends on the purposes of the evaluation.

# Formative Evaluation

- Provides information to improve programs, clarifying program strengths and limitations
- Monitors progress toward goals
- Helps refine data collection activities

# Formative Evaluation

- *Project staff* can observe and talk as a group about what is and isn't working well.
- Ditto for querying teacher participants (and resisters, and drop-outs).
- Or data can be collected by others outside of the project.

# Summative Evaluation

- Assesses progress toward program goals
- Describes the effects of the program on participants
- Quantifies the amount of change experienced by program participants


# Internal Evaluation

- Less expensive
- People collecting the data likely to have a deep understanding of the project goals and activities.

# External Evaluation

- Someone external to the program may be more credible
- External evaluator may have particular kinds of expertise

- Regardless of whether there is an external and/or internal evaluation, all evaluations start with clear goals and an understanding of what will count as evidence of quality/progress.



There are always options:  
designing evaluations

*horizon*  
RESEARCH, INC.

---

# Project Description

- Handout provides an overall description of a context with a detailed description of how the project is structured to address teacher quality goals

# Evaluation Plans

- Evaluation Plans A and B provide two options for assessing the quality and impact of a component of the MSP

# Task

- Individually, read the project description and the two evaluation plans and jot down your thoughts on the strengths and weaknesses of each evaluation plan (15 minutes)
- As a table group, discuss the strengths and weaknesses of each plan (20 minutes)

## The intervention plan:

- Is based on evidence of teachers' needs
- Uses evidence of prior effectiveness in choosing professional development materials
- Uses evidence of the appeal of the coaching to teachers
- Is phased in a way that facilitates obtaining solid evidence of impact (or lack of impact)

## Evaluation Plan A...

- Extensive focus on monitoring the quality of the professional development approach—using evidence collected through observations, surveys, teacher logs, and interviews
- A lot of evidence to be collected by evaluation team; will be very costly to collect, analyze, and report.
- Observation data collected by project staff provide additional evidence of quality
- Plan to use evidence of quality for mid-course corrections

## Evaluation Plan A...


- Evidence of impact on teachers is limited to self-report data
- Potential for bias because participants are volunteers; can't generalize results to entire population
- Plan for assessing student outcomes will not provide strong evidence of impact, although it may provide some indication of system-wide progress **if** compared to comparable districts in the state

## Evaluation Plan B...


- Minimal attention to collecting evidence on the quality of the interventions
- Includes a strong evaluation design to see if there is evidence of teacher content/pedagogical content knowledge gains
- Includes a strong evaluation design to see if there is evidence of gains in student achievement
- Plan to use impact evidence for mid-course corrections
- Again, can't generalize impact results to entire population

# The bottom line:

- There are never enough resources to examine both the quality and impact of all aspects of a complex project
- There are always trade-offs between formative and summative evaluation
- Project staff have a responsibility to monitor the quality of project activities, and to use the evidence they and the evaluators collect to make mid-course corrections.
- An evaluation needs to be selective about which aspects of a program to look at in-depth and where a more cursory examination may suffice



There are only a few ways of collecting data, each with its own set of advantages and disadvantages.

- 
- Questionnaires/Logs
  - Observations
  - Interviews
  - Assessments



# Evaluation Tools: What's Available

# LSC Teacher Questionnaire

- Has scales for:
  - Teacher Attitudes Toward Teaching;
  - Perceptions of Pedagogical Preparedness;
  - Perceptions of Content Preparedness;
  - Classroom Culture;
  - Use of Traditional Teaching Practices;
  - Use of Investigative Teaching Practices; and
  - Perceptions of Principal Support.

Banilower, E. R., Boyd, S. E., Pasley, J. D., & Weiss, I. R. (2006). Lessons from a decade of mathematics and science reform: A capstone report for the local systemic change through teacher enhancement initiative. Chapel Hill, NC: Horizon Research, Inc.

# LSC Principal Questionnaire

- Has scales for:
  - Opinions about Mathematics Instruction;
  - Factors affecting Mathematics Instruction in their School; and
  - Perceptions of their School's Progress in Implementing Targeted Reforms.

# Questionnaires

- Strengths:
  - Inexpensive to administer to large numbers of participants.
  - Reasonably valid and reliable measures of classroom practice.

# Questionnaires

- Weaknesses:
  - Unless you get high response rates, results are not meaningful.
  - Self report
    - Likely won't be valid in "high stakes" situations.
    - Just because teachers are doing more of something doesn't mean they are doing it well.

# LSC Teacher Interview Protocol

- Collects data on teachers' views of:
  - Impacts of LSC on Instruction;
  - Most Helpful Aspects of the LSC;
  - Least Helpful Aspects of the LSC;
  - Additional Help Needed;
  - Supportive Policies/Practices; and
  - Limiting Policies/Practices.

# LSC Teacher Interview Protocol

- Strengths:
  - Provides information that can be used for mid-course corrections.
  - Provides “testimonials” that can be used for gaining stakeholder support.
  - Can be easily adapted for any professional development program
- Weakness:
  - Self-report data.

# LSC District Policy Forms

- Collects data on:
  - How district policies and practices impact the process of mathematics reform in the LSC district;
  - Extent of support (or opposition) of various stakeholders toward mathematics reform in the district; and
  - Status of capacity, infrastructure, and resources devoted to mathematics reform.

# LSC District Policy Forms

- As measures, these were untested, but they had the benefit of drawing attention to the importance of an aligned system.
- Savvy PIs were able to use these data as leverage with superintendents.

# LSC Professional Development Observation Protocol (PDOP)

- Evaluator rates PD session on a number of indicators related to its:
  - Design;
  - Implementation;
  - Content (content and/or pedagogical); and
  - Culture.
- Evaluator provides rating of overall quality of PD session.

# LSC Professional Development Observation Protocol

- Strengths:
  - Provides “face validity” to evaluation.
  - Having an observation protocol helps ensure that all observers are attending to the same things.
  - Provides insight into needed improvements.

# LSC Professional Development Observation Protocol

- Considerable constraints:
  - Need to train observers to make sure they all focus on the same aspects of the PD/rate the PD consistently.
  - Observers need knowledge of the content/PCK, an understanding of teaching, and an understanding of strategies for working with adult learners.

# LSC Professional Development Observation Protocol (PDOP)

- “Bonus” of PDOP:
  - Projects were able to use the PDOP to establish a shared vision of effective professional development and as a basis for professional development for PD providers.

# LSC Classroom Observation Protocol (COP)

- Evaluator rates lesson:
  - Design;
  - Implementation;
  - Content; and
  - Culture.
- Evaluator provides rating of overall lesson quality.

# Strengths of the COP

- Provides insight into the quality of instruction.
- Provides “face validity” to evaluation.
- Having an observation protocol helps ensure that all observers are attending to the same things.

# Weaknesses of the COP

- Potential for selection bias:
  - Teachers who agree to be observed may not be representative.
  - Teachers may plan different lessons for the observed class periods, so may not be representative of their teaching.

# Weaknesses of the COP

- Observers need both knowledge of the content and an understanding of teaching.
- Need to train observers to obtain interrater reliability.
- Much more costly than surveys.
- Observation of a single lesson leaves many questions unanswered.
- Protocol is neither content-specific, nor materials-specific.

# LSC Classroom Observation Protocol (COP)


- “Bonus” of COP:
  - Projects used with teachers to establish a shared vision of effective mathematics instruction

# Identifying the “problem”

- Classroom observations provide information about impact on teaching
- But they are expensive

- And if we don't see evidence of impact in the classroom, we don't know where the breakdown occurred:
  - Lack of teacher knowledge?
  - Inability to apply knowledge to the classroom?
  - Disincentives, e.g., lack of principal support?

- What you do to improve the situation requires understanding the nature of the problems.

- 
- LSC professional development protocol, classroom observation protocol, teacher questionnaires, teacher interviews, and policy inventories.

<http://www.horizon-research.com/instruments/>

# Looking Inside the Black Box: Teacher Assessments

- Measures of teacher content knowledge are being developed
  - Ball, D. L., Hill, H.C., Rowan, B., Schilling, S. (2002) Measuring teachers' content knowledge for teaching: Elementary mathematics release items, 2002. Ann Arbor, Michigan: Study of Instructional Improvement.

# Assessments of Content Knowledge for Teaching


- Strengths:
  - Relatively objective measures of teacher content knowledge and application of content knowledge
  - Items couched in the context of teaching; more like what teachers are expected to do.
- Weaknesses:
  - Only exist (or planned) for a small number of content areas/grade-ranges.

# Embedded Assessments of Content Knowledge for Teaching


- Embedded assessments of content knowledge for teaching are being developed
  - Seago, N., Mumme, J., & Branca, N. (2004) Learning and teaching linear functions: Video cases for mathematics professional development, 6-10. Portsmouth, NH: Heinemann.

# Embedded Assessments of Content Knowledge for Teaching

- Strengths
  - Teachers see the relevance to their work
  - Opportunity to assess current knowledge in the context of professional development
- Weaknesses
  - Difficult and expensive to score reliably enough to be considered evidence of impact



An evaluation design combines some or all of the data collection strategies based on available instruments/resources to best answer the evaluation questions.



Appropriate instruments are necessary but not sufficient. In evaluating professional development you want to be able to determine if the outcomes are due to the professional development program or some other factor(s).

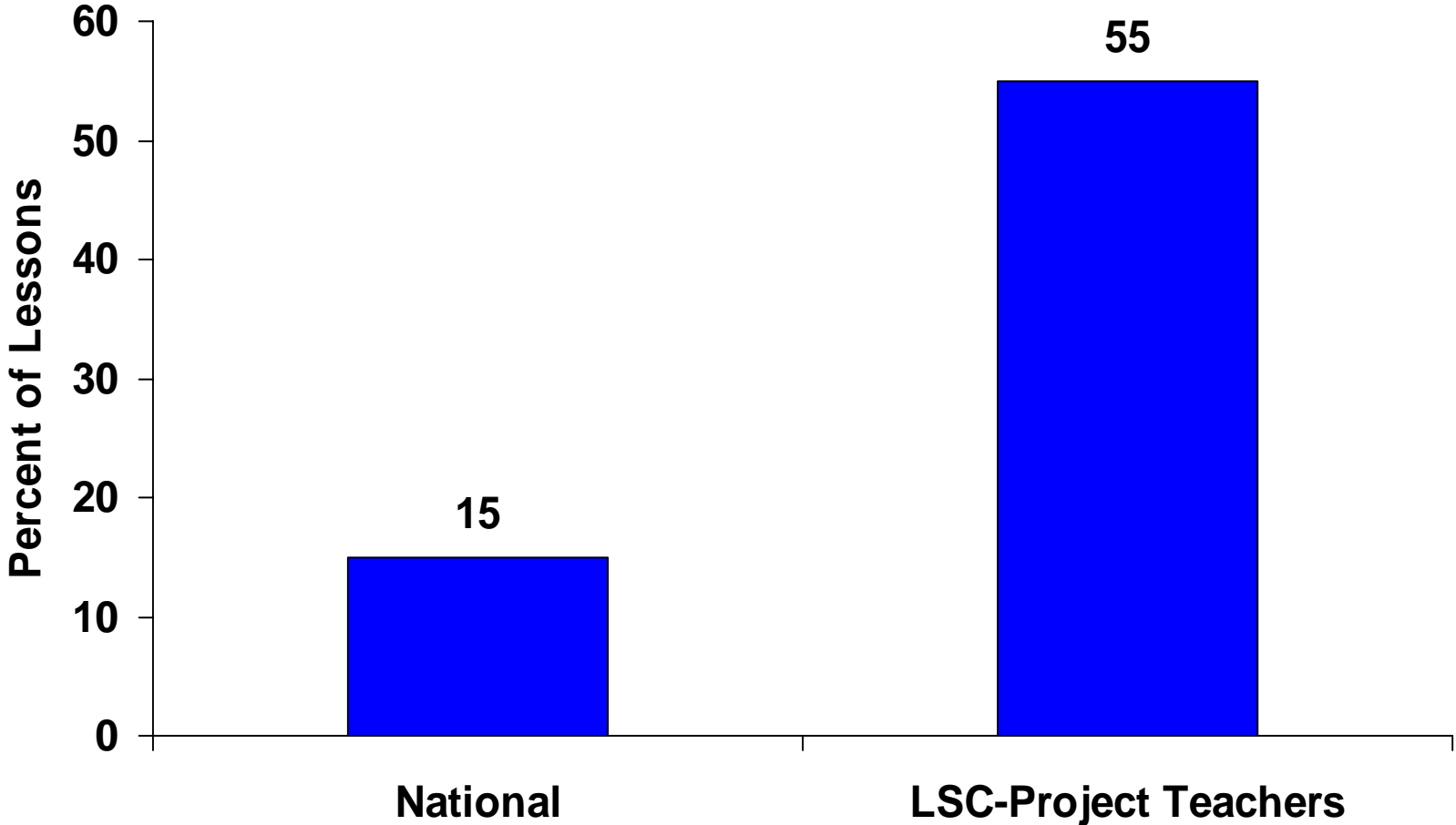



Take a look at the following data.

*horizon*  
RESEARCH, INC.

---

# Percent of K-12 Mathematics Lessons Rated Highly





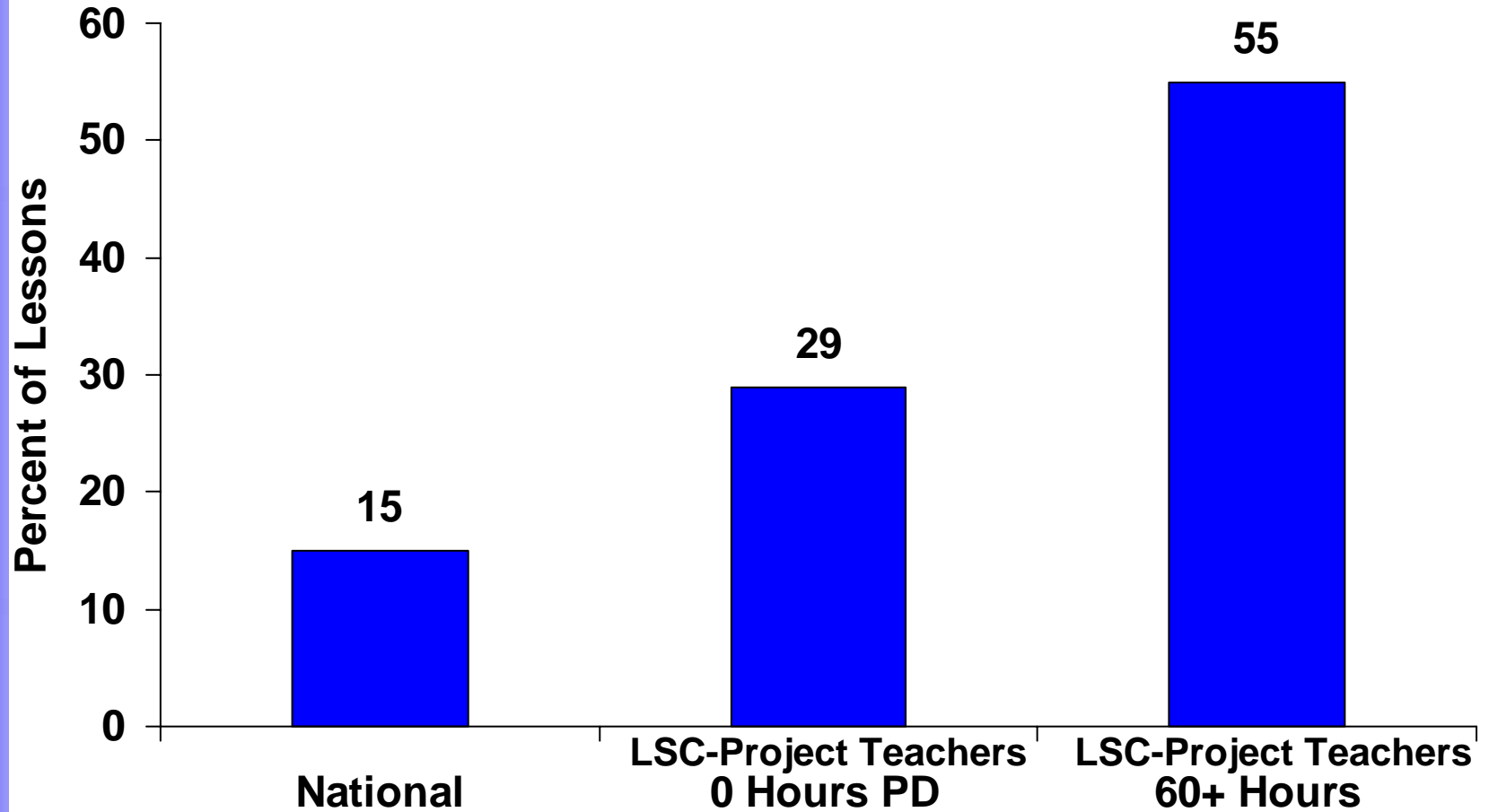
Based on this graph: Do these data convince you that the LSC was effective in improving the quality of K-12 mathematics instruction? Why or why not?

– Talk with a neighbor

# Reason for skepticism

The LSC professional development participants may have had better mathematics instruction to begin with.

## Percent of K-12 Mathematics Lessons Rated Highly



- Need to be careful producers of evidence and critical consumers of evidence.

# Avoiding Common Pitfalls

- Sampling: Choose samples that are appropriate for answering the evaluation questions
  - Results achieved with a hand-picked or volunteer group may provide evidence that a program *can* produce intended outcomes, but not of its likelihood to produce those outcomes with a larger population

- Use samples large enough to support the planned analyses, particularly for disaggregating results

- Data collection: Protect against selection/response bias
  - Self-selection into or out of interventions
  - Self-selection into or out of measurement

- Attributing impacts: Choose a design and methods of analysis that connect evidence of outcomes to the interventions

- Random field trials are “gold standard” for causality claims, but often not feasible.
- An equivalent comparison group is the next best option.
- If no comparison group is available, a variation in exposure to interventions over time and across groups can help to tease out contributions of the interventions to the outcomes.

# Reporting Progress of Reform

Evidence should be appropriately reported in terms of:

- Evidence that the professional development program produced the intended outcomes
- Educational significance as well as statistical significance
- Evidence that the interventions are being “scaled-up,” including attention to possible loss of quality or impact
- Evidence that the interventions are “going to scale,” including attention to possible loss of quality or impact

# Taking Stock

---

A Practical Guide to  
Evaluating Your Own Programs

---

by  
Sally L. Bond  
Sally E. Boyd  
Kathleen A. Rapp

with  
Jacqueline B. Raphael  
Beverly A. Sizemore

1997

**Horizon Research, Inc.**  
111 Cloister Court - Suite 220  
Chapel Hill, NC 27514-2296

[http://www.horizon-research.com/reports/1997/taking\\_stock.php](http://www.horizon-research.com/reports/1997/taking_stock.php)