Module 4

Design and Measurement Strategies

Learning Objectives

- To understand different strategies used to answer M&E questions
- To become familiar with specific designs used to answer impact and effectiveness questions
- To be familiar with common indicators and measures
- To be familiar with concept of operational definitions
- To be aware of measurement issues

Evaluation Research Design

Type of Questions Determine Strategy

1. For descriptive and normative questions, strategy may be simple and straightforward.
2. For impact questions, how you will control for other factors so you can ascertain the effects of your program?
3. Impact questions require different types of design.

Monitoring

Monitoring tends to ask descriptive questions and/or target questions

- Descriptive questions: snap shot
  - Who, what, when, where, why, how and how many?
- Normative questions:
  - Has the target been reached?

Monitoring

- Monitoring is the periodic assessment of a program/project to determine its progress toward intermediate or final objectives.

- Validity of indicators for monitoring is crucial: Are you measuring what you think you are measuring? Are you measuring what counts?

Challenge of Impact Questions

To demonstrate impact you must be able to eliminate other possible explanations.

Internal Validity

Goal of Design Strategy for impact questions is Internal Validity

Eliminates other possible explanations

Internal Validity

Internal Validity refers to the extent to which the design enables you to determine that the program, rather than other factors, caused the changes you have observed.

This is important when answering *impact* questions.

Threats to Internal Validity

- There are several threats (examples)
- Does not mean they actually exist.
- You want to consider the plausibility.

Importance of Design

- Designs attempt to eliminate or reduce other possible explanations.
- Design is crucial in evaluations that want to show that the program caused the desired result or had an impact.
- Design is not important in monitoring.

General Types of Designs for Answering Impact Questions

- Experimental
- Quasi-Experimental
- Non-Experimental

Types of Design

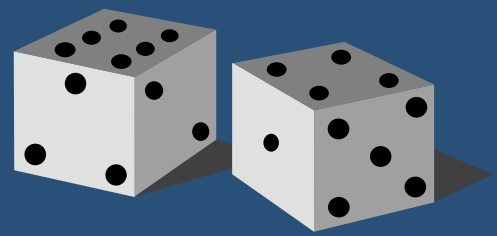
Experimental Design

Key elements:

- **Random Assignment**
- Comparison (with and without the Program)
- Before and After Measurement

Experimental Design

R	O_1	X	O_2
R	O_1		O_2



R indicates **R**andom assignment

O is the **O**bservation or measure

X is the Program (or treatment)

Types of Design

Quasi-Experiment

"Quasi" means no random assignment

Key elements:

- Comparison (with and without the Program)
- Might include before and after measures

Types of Design

Quasi-Experimental Designs (Cont'd)

- Use when you cannot control the process for deciding who gets the treatment.
- Weak because there may be selection bias and other biases
- But this is often more practical in public sector research

Quasi-Experimental Design

O_1 X O_2 Program Group

O_1 O_2 Control Group

- Groups
  - Matched pairs
  - Non-equivalent comparison groups.

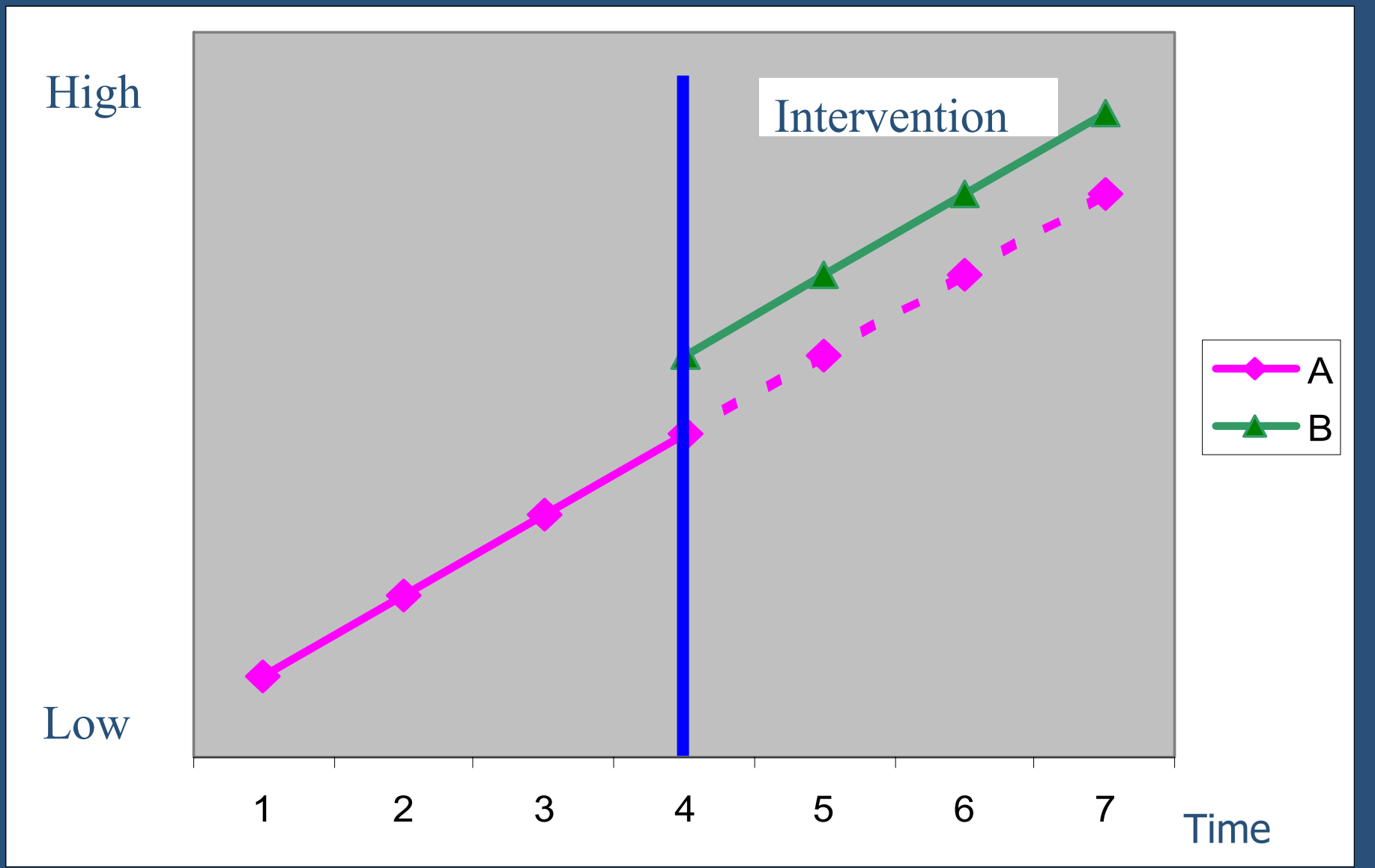
Quasi-Experimental Design: Interrupted Times Series

- Key elements: many measures before and after the "treatment"

$O_1 \quad O_2 \quad O_3 \quad X \quad O_4 \quad O_5 \quad O_6$

(Some suggest that you should have at least 10 measures before and after the treatment (X))

Interrupted Time Series Design



Quasi-Experimental Design: Interrupted Times Series

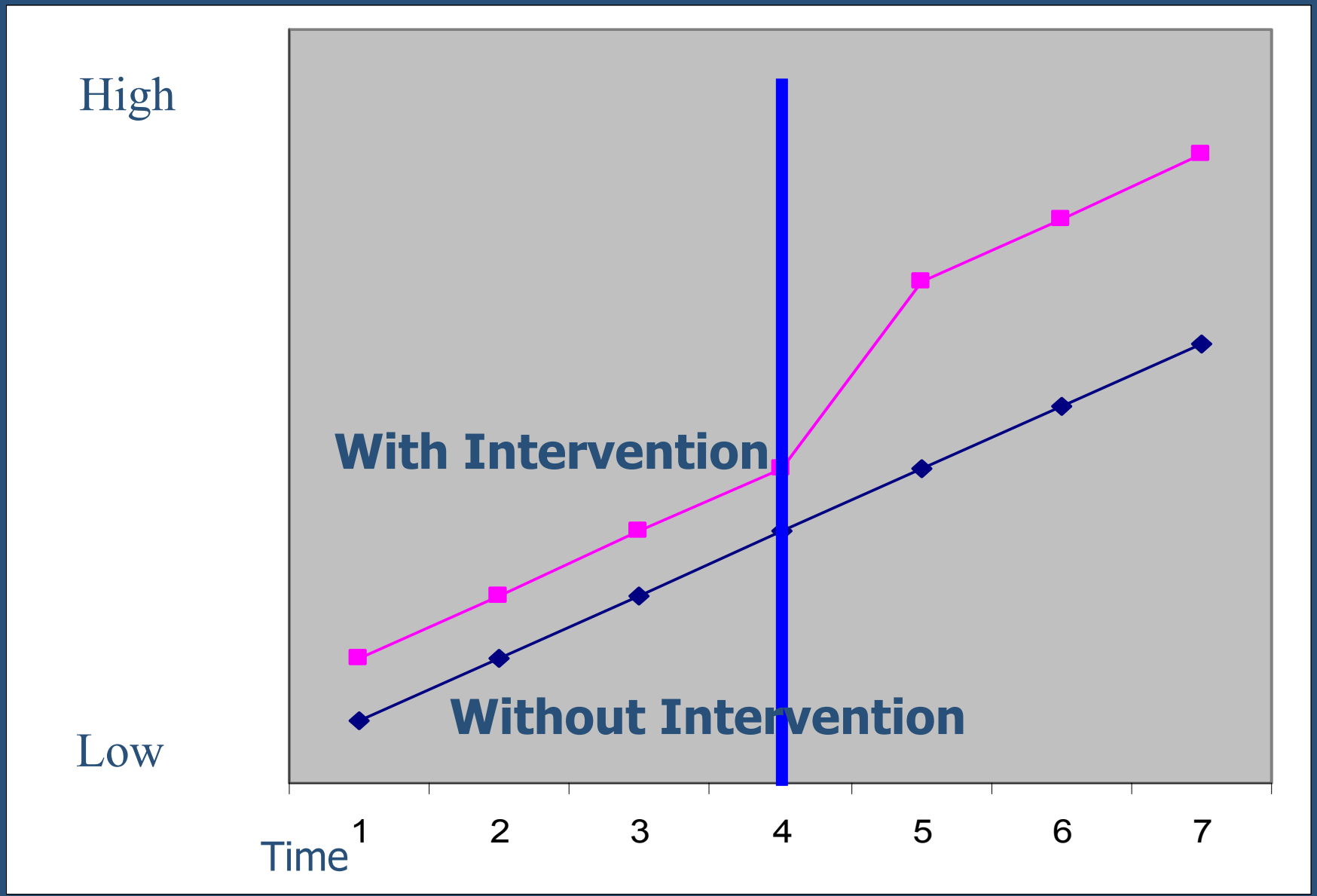
- Key elements: many measures before and after the "treatment"

O_1 O_2 O_3 X O_4 O_5 O_6

O_1 O_2 O_3 O_4 O_5 O_6

(Some suggest that you should have at least 10 measures before and after the treatment (X))

Interrupted Time Series Design



Types of Design

Non-Experimental Designs

Key elements:

- No random assignment
- Maybe no before-program measures
- Maybe no comparison

Non-Experimental Design

Before and After Design

O X O

Static Group Comparison

X O

O

One Shot: X O

Discussion: Applying Design to Case

Given what you know about the case, what type of design could it use to:

- Determine the impact on teacher attitudes?
- Determine the impact on student test scores?
- Determine whether school administrators are satisfied with the performance of teachers who participated in the training?

Measurement Strategy

- What do you want to know?
- How will you know it?
- What data do you need?

Developing Measurement Strategy

- **Conceptual definition**
  - Key terms:
    - training, innovation, attitudes
  - Boundaries:
    - in 9 districts, from 1999-2002.
- **Operational definition**
  - How it will be measured?
    - implies numbers

Indicators/Monitoring

- Monitoring program performance at repeated intervals to track progress requires the use of carefully identified and defined indicators so that meaningful comparisons can be made.
- The definition and measurement issues we discuss here are common to both monitoring and evaluation.

Definitions

- An indicator is a word or phrase which "indicates" the level or extent of some phenomenon of interest

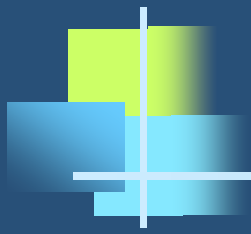
  (Case: knowledge of math)

- A measure is the operational definition of how data are collected to assign a value to an indicator

  (Case: test score on national math exam)

Defining Your Terms

- It means translating vague words into specific meanings.
- Defining your terms means obtaining agreement from the stakeholders about the question, the definitions and the measurement strategy.



Defining Your Terms

- Sometimes it is difficult to assign a number or to actually measure what you want to measure.
- For example, you may not really be able to measure the quality of a program. Instead, you may have to be content with measuring whether people think it is a quality program.

Examples

- Teacher attitudes:
  - Measured by using a survey that asked teachers about their attitudes.
- Quality of instruction:
  - Measured by having observers rate specific components of performance.
- Effectiveness of the training system:
  - Measured by the number of participants.
  - Measured by meeting set targets for teacher attitudes and innovations in instruction.

Case: Measures

1. Did teacher attitudes change after the training?
   a. Indicator: attitudes
   b. Measure: responses to a series of attitude questions (0-4 scale)
2. Did students improve their math and science knowledge?
   a. Indicator: knowledge
   b. Measure: test scores

Some Commonly Used Measures

- Frequencies, percents
- Means, Medians, Modes
- Dollars, Euros, Yen
- Percent change
- Rates, Ratios

Key Issues about Measures

- Are they relevant?
- Are they valid?
- Are they reliable?
- Are they precise?

Data Source Issue

- Where are the best sources of data?
  - Validity
  - Reliability
- Do the data already exist?
  - Are they reliable?
- Do you have to collect new data?

Discussion: Where Can We Find Data?

- Number training seminars held
- Number of teachers who completed training
- Attitudes of teachers
- Quality of teaching
- Quality of teaching material
- Participation of students
- Knowledge of students

Data Source Lessons

- Which ones might be easier to obtain?
- Which ones might be very difficult to obtain?
- How accurate and reliable are each of the data sources?
- How valid are existing data?
- What other concerns might you have?

Case Discussion

- Goal: Capability of young people in Mathematics and Science is upgraded.
  - How do they define capability?
  - What strategy did they use?

Evaluation Grid

- One tool that some find useful is the evaluation grid
- This tool helps you see how you intend to answer each question
- Use the evaluation grid in your workbook
- For each question, you will need to identify the information needed, sources of that information, and how you will collect the data
- This is an iterative process

PDM/Evaluation Grid

- PDM helps you clarify what to evaluate.
  - What are inputs, activities, outputs and intended outcomes?
  - What are causal relationship among project components?
  - What are indicators they are using?
  - What are risks for the project?
- Evaluation Grid:
  - Identifies the questions, data needed and strategy for conducting the evaluation

Exercise 3: Evaluation Grid

- Return to the questions from Exercise 2. Select 3 questions and place them on the evaluation grid.
- For each of your questions:
  - What data/measures would best answer your questions?
  - What are likely sources of information?
- Complete the
  - Data Needed and Measures column
  - Source of Information column

Discussion

Discuss Exercise 3 Evaluation Grid (Design Matrix)