

Measurement Issues in Evaluation

1

Evaluation Designs

2

Selecting an Evaluation Design

- How much time do you have?
- What financial resources are available?
- How many participants can be included?
- Qualitative or quantitative data (or both)?
- Do you have data analysis skills?
- Do you want to generalize your findings to other populations?

3

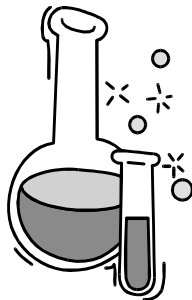
Selecting an Evaluation Design

- Are stakeholders concerned with validity and reliability?
- Can you randomize participants into experimental and control groups?
- Do you have access to a comparison group?

4

Experimental Approaches

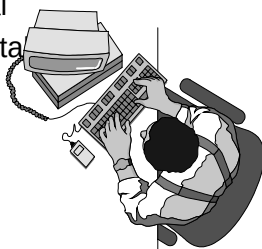
- Traditional type of research
- Only type of research performed in the sciences
- Future-focused
- Involves manipulation of the independent variable(s)



5

Types of Experimental Designs

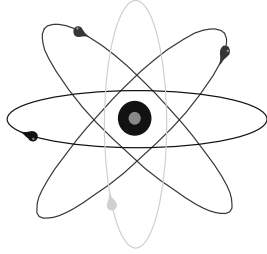
- Non-experimental
- Quasi-experimental
- Experimental



6

Research Design Notation

- (R) = Random assignment
- X = Treatment
- O = Observation



7

Non-Experimental

- Primitive designs
- Designed only to show the effects of a treatment
- Do not control for internal validity



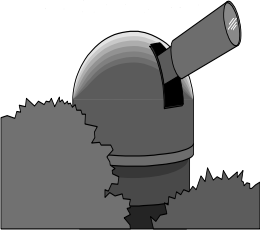
8

Non-Experimental

- No random assignment of subjects
 - Pretest-posttest design
 - Test administered, followed by the program, then the same test administered
 $O_1 \quad X \quad O_2$
 - Time series design
 - A series of observations to establish baseline, followed by the program then test administered
 $O_1 \quad O_2 \quad O_3 \quad X \quad O_4 \quad O_5 \quad O_6$

9

Quasi-Experimental



- Better than non-experimental
- When subjects cannot be randomly selected or randomly assigned to treatment groups
- Controls some threats to internal validity

10

Quasi-Experimental Designs

- Promote external validity & maintain internal validity

1. Pretest-posttest design

O₁ X O₂ (experimental)
O₃ O₄ (control)

2. Time series design

O₁ O₂ O₃ X O₄ O₅ O₆ (experimental)
O₇ O₈ O₉ O₁₀ O₁₁ O₁₂ (control)

11

Experimental



- Recommended
- Random selection of subjects from the population
- Random assignment of subjects to treatment groups
- Controls threats to internal and external validity

12

Experimental Designs

■ Better control of research short-comings

□ Pretest-posttest design

- Random assignment to groups, test all groups, program to some groups then test all groups again

(R)	O ₁	X	O ₂ (experimental)
(R)	O ₃		O ₄ (control)

□ Posttest only design

- Random assignment, then program to some groups, then test all groups

(R)	X ₁	O ₁ (experimental)
(R)	X ₂	O ₂ (experimental)
(R)		O ₃ (control)

13

Experimental Designs

3. Time series design

(R)	O ₁	O ₂	O ₃	X	O ₄	O ₅	O ₆
(R)	O ₇	O ₈	O ₉		O ₁₀	O ₁₁	O ₁₂

14

There are many different possible designs...

Figure 14.4 Evaluation designs

I. Experimental design									
1. Pretest-posttest design	(R)	O ₁	X	O ₂					
—Experimental group	(R)	O ₁		O ₂					
—Control group									
2. Posttest-only design	(R)		X	O					
—Experimental group	(R)			O					
—Control group									
3. Time series design	(R)	O ₁	O ₂	O ₃	X	O ₄	O ₅	O ₆	
—Experimental group	(R)	O ₁	O ₂	O ₃		O ₄	O ₅	O ₆	
—Control group									
II. Quasi-experimental design									
1. Pretest-posttest design		O ₁	X	O ₂					
—Experimental group		O ₁		O ₂					
—Comparison group									
2. Time series design		O ₁	O ₂	O ₃	X	O ₄	O ₅	O ₆	
—Experimental group		O ₁	O ₂	O ₃		O ₄	O ₅	O ₆	
—Comparison group									
III. Nonequivalent design									
1. Pretest-posttest design		O ₁	X	O ₂					
—Experimental group		O ₁		O ₂					
—Comparison group									
2. Time series design		O ₁	O ₂	O ₃	X	O ₄	O ₅	O ₆	
—Experimental group		O ₁	O ₂	O ₃		O ₄	O ₅	O ₆	
—Comparison group									

Key: (R) = Random assignment
O = Measurement/Observation
X = Program/Intervention

15

Internal Validity

16

Internal Validity

- The degree to which the program caused the change that was measured?
- In other words, were the changes in the participants due to your program or something else?



17

Threats to Internal Validity

- | | |
|--------------------------|---|
| ■ History | ■ Experimental Mortality |
| ■ Maturation | ■ Diffusion or imitation of treatments |
| ■ Testing | ■ Compensatory equalization of treatments |
| ■ Instrumentation | ■ Compensatory rivalry |
| ■ Statistical Regression | |
| ■ Selection | |

18

Threats to Internal Validity

- History:
 - Something happens during the program to influence the results (e.g. During your program on healthy eating, the low carb phenomenon hits)
- Maturation:
 - People may change because they grow older, wiser, stronger
- Testing:
 - Pre-test/Post-test results are different because people are familiar with the tests used

19

Threats to Internal Validity

- Instrumentation:
 - Change in measurement procedures (e.g. M-F measures of caloric intake vs. M-Su)
- Statistical Regression:
 - Extremely high or low scores on the pre-test will come back to the center on the post-test
- Selection:
 - Differences between the experimental and control groups due to lack of randomization

20

Threats to Internal Validity

- Mortality/Participant Attrition:
 - Participant drop-out. Who is left standing at the end? Has the make-up of the group changed significantly?
- Diffusion or Imitation of Treatments:
 - Control group interacts with experimental group and imitates their behavior
- Compensatory Equalization:
 - It's not fair!! Control group whines about not getting the treatment or program

21

Threats to Internal Validity

- **Compensatory Rivalry:**
 - When the control group is seen as the “underdog” and tries harder on the post-test (a.k.a. Avis Effect)
- **Resentful Demoralization (more whining):**
 - One treatment group may perceive another group to be getting a better treatment or program so they don't try as hard or give up

22

Methods of Control

Randomization (controls many threats to internal validity)

- Random selection of participants from the target population
- Random assignment to treatment or control groups (experimental)
- Random assignment of types of treatment or no treatment to groups (quasi-experimental)

23

Measurement Validity

24

Measurement Validity

■ Validity

- The degree to which an instrument measures what the evaluator wants it to measure
- Types:
 - Face
 - Content
 - Criterion
 - Construct



25

Measurement Validity

■ Face validity

- Extent to which the instrument appears to be measuring what it is supposed to measure

■ Content validity

- Extent to which an instrument samples items from the universe of content desired

26

Measurement Validity

■ Criterion validity

- Extent to which an instrument correlates with another more accurate instrument (the criterion)

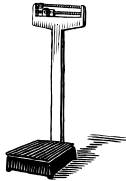
■ Construct validity

- Extent to which the measure correlates with other measures in predicted ways, but for which no true criterion exists

27

Reliability

- An empirical estimate of the extent to which an instrument produces the same result, applied once or two or more times



28

Biases to Valid and Reliable Measurement

- Guinea pig effect
 - People who are aware that they are being measured may respond in uncharacteristic ways
- Role selection
 - Awareness of being measured may influence people to feel that they have to play a special role

29

Biases to Valid and Reliable Measurement

- Measurement as a change agent
 - The act of taking a measurement may affect the subsequent behavior or those being measured.
- Response sets
 - People respond to questionnaires and interview in a socially desirable, predictable way that have little or nothing to do with the questions posed.

30

Biases to Valid and Reliable Measurement

- Interviewer effects
 - Characteristics of the interviewer may affect respondents' receptivity and answers.
- Changes in the instrument
 - If an instrument is used more than once, a learning effect is possible.
 - Interviewers may become more proficient at or tired of implementing an interview schedule the longer they use it.

31

Biases to Valid and Reliable Measurement

- Population restrictions
 - The method of data collection may impose restrictions on the population to which the results can be generalized.
- Population stability over time
 - An instrument administered at different times may not collect the same data on different populations

32

Biases to Valid and Reliable Measurement

- Population stability over areas
 - The same way of collecting data in two different geographic areas may assess different types of people
- Content restrictions
 - A limited range of data can be reported by each methods

33

Biases to Valid and Reliable Measurement

- Stability of content over time
 - If a program restricts a study to naturally occurring behavior, the content of the studied phenomenon may vary over time.
- Stability of content over an area
 - A program may be uniform in content throughout the area in which it is applied.

34

Data Collection Methods

35

Types of Variables

- Demographic
- Cognitive
- Skill
- Psychosocial
- Behavioral
- Environmental
- Health
- Cost



36

Obtrusive Measures

- Self-completion questionnaires
- Self-completion main surveys
- Self-completion diaries and logs
- Face-to-face interviewing
- Telephone interviewing
- Direct observation

37

Unobtrusive Measures

- Abstraction of existing medical and clinical records
- Clinical or physiological measures



38

External Validity

39

External Validity

- Can the findings be generalized beyond the subjects or settings studied in the experiment?
- If we do this again with some other group, will it work again?



40

Threats to External Validity

- Social desirability
 - Subject responds the way he/she thinks the experimenter wants them to
- Expectancy effect
 - When people act as you expect them to act (Rosenthal effect)

41

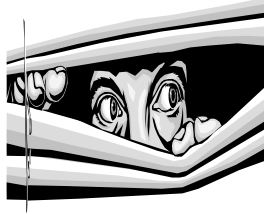
Threats to External Validity

- Hawthorne effect
 - Behavior change due to people feeling special
- Placebo effect
 - Change in behavior due to belief in the effectiveness of the treatment

42

Controlling Threats to External Validity

- Blind
 - * Participants
- Double blind
 - * Participants and investigators
- Triple blind
 - * Participants, investigators, evaluators



43
