Designing Quantitative Research

Katie Rommel-Esham
Education 504

Threats to Validity

Factors other than the independent variable which provide plausible rival hypotheses (PRH) to the treatment effect

Internal Validity

Asks the question “Did the experimental treatment in fact make a difference in this specific instance?”
Threats to Internal Validity

**History**
- Specific events (in addition to the experimental variable) that occur between the first and second measurement
- Includes things like different teachers, different time of day, local events, TV shows

Threats to Internal Validity (con’t)

**Selection**
- Artifact of different kinds of respondents in comparison groups
- May be controlled using randomization

Threats to Internal Validity (con’t)

**Maturation**
- Processes within the respondents operating as a function of time (between pretest and posttest)
- Includes growing more “something” such as older, wiser, stronger, more experienced, hungrier, more tired, etc.
**Threats to Internal Validity (con’t)**

- **Reactive or Interaction Effect (of testing)**
  - Pretest may increase or decrease respondents’ sensitivity to the experimental variable
  - The effects of taking a test on the scores of a second testing (the number of times particular responses are measured)
  - Can never really “erase” prior knowledge gained by completing an instrument at an earlier time

**Threats to Internal Validity (con’t)**

- **Instrumentation**
  - Measurement errors that result from changes in the calibration of an instrument or changes in the observers, scorers, or the instrument itself
  - Inter-rater reliability plays a significant role here

**Threats to Internal Validity (con’t)**

- **Treatment Replications**
  - If a treatment is administered to a group, this counts as one administration of the treatment, not \(n\) administrations, where \(n\) is the number of individuals in the group
  - Thus reported results can be misleading if each subject in a group administration is counted as an individual replication
Threats to Internal Validity (con’t)

- **Experimental Mortality**
  - Differential rates of loss from comparison groups
  - Also deals with subject attrition for any particular group involved, not exclusively a differential loss
  - A particular problem with longitudinal studies

Threats to Internal Validity (con’t)

- **Statistical Regression**
  - If groups have been chosen on the basis of extreme scores, then regression toward the mean is likely to occur
  - This is a result of measurement error
  - For example, students with extremely high scores will tend to receive lower scores in a subsequent testing, as they have no other direction in which to go

Threats to Internal Validity (con’t)

- **Diffusion of Treatment**
  - Members of different groups who come in contact with each other cause the treatment to diffuse
  - Those intended to be in the control group may interact with those in the treatment group in such a way that the treatment is then spread to the control group
Threats to Internal Validity (con’t)

**Experimenter Effects**
- Attributes or expectations of the researcher, either deliberate or unintentional, that influence the subjects
- May be differential treatment (tone of voice, reinforcing different behaviors, being more reassuring to one group, displaying different attitudes), or characteristics that affect responses (age, clothing, gender, educational level, race)

Threats to Internal Validity (con’t)

**Subject Effects**
- Changes in the subjects that result from their awareness of being subjects
- Includes
  - Hawthorne Effect (an increase in desirable behavior),
  - John Henry Effect or Compensatory Rivalry (where subjects try harder because they see themselves in competition with the treatment group),
  - Resentful Demoralization (subjects become unmotivated when they are not selected for the “preferred” treatment), and
  - Novelty Effect (subjects react positively because they are doing something new and different)

Threats to Internal Validity (con’t)

**Interactions with Selection:**
- Selection-Maturation Interaction
- Selection-History Interaction
- Selection-Testing Interaction
- Effects resulting from an interaction between the way the comparison groups were selected and their maturation, history events, and/or testing effects over time
Threats to Internal Validity (con’t)

- Ambiguity About the Direction of Causal Influence
  - Occurs when it is not clear whether A causes B, or B causes A

External Validity

- Asks the question “To what populations, settings, treatment variables, and measurement variables can this effect be generalized?”

Threats to External Validity (con’t)

- Interaction of Selection and Experimental Variable
  - Some groups may be more affected by the treatment because of the composition of the group
  - Becomes more likely as getting subjects becomes more difficult
Threats to External Validity (con’t)

- Reactive Effects of Experimental Arrangements
  - Factors which would preclude generalization to those exposed to the treatment outside of the non-experimental settings
  - A good example is a residential stuttering clinic

- Multiple-Treatment Inference
  - Occurs when multiple treatments are applied to the same group of respondents because effects of prior treatments are not generally “erasable”

- Interaction of History and Treatment
  - Problematic when an experimental situation takes place on a particularly memorable day (for example on 9-11)
  - Would the same result be observed under more “mundane” circumstances?
Quasi-Experimental and Experimental Research Designs

Pre-Experimental Designs (con't)

**Single-Group Pretest-Posttest Design**
- Threats to validity increase as time increases and experimental situations become less controlled and more contrived

```
O → X → O
Pretest Treatment Posttest
```

Pre-Experimental Designs (con’t)

**Non-Equivalent Groups Posttest-Only Design**
- Lack of pretest cannot allow us to rule out selection as a plausible rival hypothesis

```
A → X → O
Treatment

B → O
Posttest
```
Pre-Experimental Designs (con’t)

- **Non-Equivalent Groups Posttest-Only Design** and **Multiple-Groups Multiple Treatments Posttest Only Design**
  - Lack of pretest cannot allow us to rule out selection as a plausible rival hypothesis
  - May be extended to as many groups as needed
    
    \[ \begin{align*}
    \text{A} & \quad X_1 \quad O \\
    \text{B} & \quad X_1 \quad O \\
    \text{C} & \quad X_1 \quad O \\
    \text{Treatment} & \quad \text{Posttest}
    \end{align*} \]

Quasi-Experimental Designs (con’t)

- **Nonequivalent-Groups Pretest-Posttest Design** and **Multiple-Groups Multiple-Treatments Pretest-Posttest Design**
  - Often used for intact or pre-existing groups like classrooms
  - May be extended to as many groups as needed
    
    \[ \begin{align*}
    \text{A} & \quad O \quad X_1 \quad O \\
    \text{B} & \quad O \quad X_2 \quad O \\
    \text{Pretest} & \quad \text{Treatment} & \quad \text{Posttest}
    \end{align*} \]

True Experimental Designs (con’t)

- **Randomized-Groups Posttest-Only Design** and **Randomized-Groups Multiple-Treatments Posttest-Only Design**
  - Randomization helps to control selection as a plausible rival hypothesis
  - May be extended to as many groups as needed
    
    \[ \begin{align*}
    \text{A} & \quad X_1 \quad O \\
    \text{B} & \quad X_2 \quad O \\
    \text{Treatment} & \quad \text{Posttest}
    \end{align*} \]
True Experimental Designs (con’t)

- Randomized-Groups Pretest-Posttest Design
  - May be extended to as many groups as needed

```
R  A  O  X1  O  \\
  B  O  X2  O
Pretest  Treatment  Posttest
```

Factorial Designs

- Used to look for interaction between two or more independent variables
- May be experimental or nonexperimental

```
Females
R  A  O  X1  O  \\
  B  O  X2  O
Pretest  Treatment  Posttest
```

```
Males
R  A  O  X1  O  \\
  B  O  X2  O
Pretest  Treatment  Posttest
```

Single-Case Experimental Designs
(Schloss & Smith, 1998)

- Used to assess performance changes (particularly in special education)
- Objective and efficient
- Well suited to many academic and social performance problems
- Can serve as a foundation for more complex analytic methods
- Well-designed studies rule out threats to internal validity, but cannot control for them
- History and maturation are particularly relevant in this case
Single-Subjects Designs

*A-B Design*
- Most simple and least interpretable
- Observe until undesirable behavior is at a consistent, stable rate, then introduce treatment
- This is based on the premise that if no treatment were introduced, undesirable behavior would continue
- If behavior does change, it may be attributable to the intervention
- Weak in internal validity because it does not address PRH such as testing and history

Single-Subjects Designs (con’t)

*A-B Design*

<table>
<thead>
<tr>
<th>Intervention</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>XXXXXXXXX</td>
<td>XXXXXXXXX</td>
</tr>
<tr>
<td>Baseline</td>
<td>Treatment</td>
</tr>
</tbody>
</table>

A-B Design Data

![Graph showing A-B Design Data](image-url)
Single-Subjects Designs (con’t)

**Reversal (A-B-A or A-B-A-B) Design**
- Reversals (systematically introducing and removing the treatment) provide replication of treatment
- Provides a strong defense against PRH when multiple reversals are used
- Baseline data are collected before the treatment is imposed; during treatment, behavior should change in desired direction; behavior returns to baseline when treatment is removed
- Repeated demonstration of the influence of the treatment increases confidence in its effectiveness

---

**A-B-A-B Reversal Design**

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>XXXX</td>
</tr>
<tr>
<td>0000</td>
<td>0000</td>
</tr>
</tbody>
</table>

Baseline | Treatment | Baseline | Treatment

---

A-B-A-B Reversal Design Data
Single-Subjects Designs (con’t)

* Multiple-Baseline Designs
  - Employs A-B logic
  - Collection of data on two or more actions, subjects, or situations or some combination thereof
  - External validity is quite limited, however generalizability may be increased by replication with other subjects and different settings

Multiple Baseline Data