## Chapter 8. Experiments

- Experiments usually involve two stages
  - Taking action
  - Observing the consequences of that action.
- Topics covered in this chapter
  - Topics appropriate for experimental research
  - The classical experimental design
  - Variation of experimental design: Pre-experimental designs
  - Internal and external validity issues
  - Strengths and weaknesses

## Topics Appropriate for Experimental Research

- Hypothesis testing involving limited and well-defined concepts and propositions.
- Explanatory purposes rather than descriptive purposes
- Laboratory or natural experiments

## The Classical Experimental Design

- Purpose
  - Examine the effect of an experiment stimulus (the independent variable) on a dependent variable.
- Procedure
  - Step 1. Randomly assign subjects in your sample to an experimental group and a control group.
  - Step 2. Pretest two groups to make sure they are similar in ways related to your experiment.
  - Step 3. The experimental group is exposed to the experimental stimulus, but not the control group.
  - Step 4. Compare whether the experiment group behave differently from the control group after the experiment.

## An Example

- Harvard Test of Inflected Acquisition (HTIA) (Rosenthal & Jacobson, 1968)
  - To test the theory that the way others perceive us is largely conditioned by expectations they may have in advance.
  - Process
    - The HTIA test was administered to a group of students
    - Half of the subjects were randomly selected to be in the experimental group, the other half in the control group
    - The teachers were told that those students in the experimental group are very likely to exhibit a sudden spurt in academic abilities during the coming year based on their score on the test.
  - Dependent variable
    - Academic performance in the coming year (another real test)
  - Independent variable
    - Other's expectation (whether a student was described to the teachers as being likely to exhibit a sudden spurt soon)

## An example - continued

- Result:
  - In the following year, when the researchers went back to look at academic performance of the students, they found that the students in the experimental group far exceeded those in the control group
  - Because teachers gave students the grades, it stands to reason that teachers gave higher grades to those students in the experimental groups because researchers told them those students were likely to exhibit a sudden spurt in academic abilities.
- Note that the HTIA test is a fake. It was used only to build up teachers' expectations. Whether somebody is in the experimental group or not is totally random and has nothing to be with the HTIA test scores.

## What if one cannot use classical experimental design due to circumstance? - Three pre-experimental designs.

- One-shot case study
- One-group pre-test and post-test design
- Static-group comparison

## What is "One-shot case study?"

- Characteristics of one-shot case study
  - No control group, only experimental group
  - No pre-test
  - Compare the result with some intuitive standard
- An example:
  - One wants to determine whether reading to children an extra ½ hour a day would increase their reading skill.
  - A group of children are chosen. The teacher will read an extra ½ a day to these children.
  - It was found, at the end of the semester, these children's reading skills are pretty good.
  - The problem is that one does not know whether without this extra ½ hour of reading, these students can do just as well. One does not even know whether their reading skills have even improved given there is no pre-test. Thus it cannot be said that this ½ hour extra reading lead to increased reading skills for these children.

## What is "One-group pre-test and post-test design?"

- Characteristics
  - No control group, only experimental group
  - There is a pre-test
  - Compare the result with pretest
- An example:
  - One wants to determine whether reading to children an extra ½ hour a day would increase their reading skill.
  - A group of children are chosen. A test is performed to evaluate their current reading skill (pre-test).
  - The teacher will read an extra ½ a day to these children (the experimental stimulus).
  - It was found, at the end of the semester, these children's reading skills are better compared to before.
  - The problem is that one does not know whether without this extra ½ hour of reading, these students can do just as well. Maybe their regular reading assignments are good enough for this improvement. Thus it cannot be said that this ½ hour extra reading has caused this increase in reading skills for these children.

## What is "Static-group comparison?"

- Characteristics of Static-group comparison
  - Has two groups, but no experiment is administered.
  - One group participated in some activities, the other group did not.
  - Compare the results of these two groups
- An example:
  - One wants to determine whether reading to children an extra ½ hour a day would increase their reading skill.
  - A teacher offers an extra ½ reading session everyday. Those students who want to participate can participate.
  - It was found, at the end of the semester, these children who participated this extra ½ hour reading session had better reading skills than those who did not participate.
  - The problem is that one does not know whether without this extra ½ hour of reading, these students who participated can do just as well. It is possibly those with better reading skills to begin with are more likely to participate in this extra reading session. Thus it cannot be said that this ½ hour extra reading has caused this increase in reading skills for these children.

## What is internal invalidity in experiments?

- Internal invalidity is the possibility that the conclusions drawn from the experimental results may not accurately reflect what was going on in the experiment itself.

## What are some sources of internal invalidity?

- History:
  - Historical events may occur during the course of the experiment. Such events will complicate the experimental results. For example, A race riot during the course of an experiment on reducing anti-African-American prejudice is likely to affect the internal validity of the experiment.
- Maturation:
  - Subjects grow older and wiser during the course of the experiment, especially if the experiment lasts a long time.
- Instrumentation:
  - Use of different measurements for the same concept in pretest and posttest will cause problems. The intensity of the measurements maybe different and may cause subjects to answer differently.

## What are some sources of internal invalidity? - Continued

- Statistical regression
  - It becomes a problem when subjects are selected for their extreme scores on the dependent variable. For example, a help session is designed to help kids who were ranked at the bottom 10% on midterm. Chances are, even without any help, it's not likely all of them are going to stay at the exact bottom 10% for the final. Thus even if their ranks increase on the final, it does not necessarily mean the helping session worked.
- Selection bias
  - The research selects certain people to the experimental group and certain other people to the control group. Comparison do not have any meaning unless the groups are comparable.
- Experimental mortality
  - Subjects drop out of the experiment. Those who stay in the experiment may be systematically different from those who left.

## What are some sources of internal invalidity? - Continued

- Diffusion or imitation of treatments
  - Subjects in the experimental group and subjects in the control group are communicating with each other. For example, in an educational experiment, two kids, one in the experimental group and the other in the control group, are studying together and sharing information
- Compensation
  - Experimenters sometimes decide to compensate the control group. For example, in an educational experiment, the teachers are more lenient in grading students in the control group or give extra hints to students in the control group in a testing situation.
- Compensatory rivalry
  - The control group may try to work harder than usual, especially in educational experiments.
- Demoralization
  - Control group may give up, especially in educational experiments. Under such situations, the control group may stop studying or act angry because they are not getting the help the experimental group is getting.
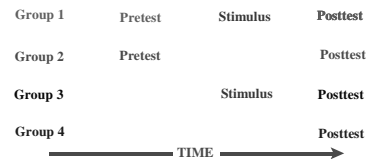
## How to improve internal validity in experiments?

- Classical experimental designs take care of the problems of internal invalidity 1-6.
- Double-blind experimental designs take care of the problems of internal invalidity 7-10.
  - In double-blind experimental designs, neither the researcher nor the subjects know who are in the experimental group and who are in the control group.

## What are external invalidity issues in experiments?

- External invalidity – Can the experimental results be generalized to the real world? – The issue of generalizability.
- External invalidity can be caused by the interaction between the testing situation and the experimental stimulus.
  - The textbook has a good example of such a situation.

## How to improve external validity? – The Solomon four-group design

| Group 1 | Pretest | Stimulus | Posttest |
| Group 2 | Pretest | | Posttest |
| Group 3 | | Stimulus | Posttest |
| Group 4 | | | Posttest |

←————— TIME ——————→

**The Solomom Four-Group Design**

① Posttest should show higher score than the pretest
② Pretest and posttest should show the same score
③ Group 1 should score higher than Group 2
④ Group 3 should score higher than Group 4

## What are the strengths and weaknesses of the experimental method?

- Strengths
  - Isolation of the experimental variable and its impact over time
  - Replication is relatively easy
- Weaknesses
  - Artificiality of laboratory experiments
  - Generalizability of results

## An article using experimental research method

- Rash, Judy, Terry D. Johnson & Norman Gleadow (1984). Acquisition and retention of written words by kindergarten children under varying learning conditions. Reading Research Quarterly, 19, 452-460.
- Theory and hypotheses:
  - H1: Samuel's focus attention hypothesis: learning in isolation is better
  - H2: Goodman's contextual information important: learning in context is better

- Measurements:
  - Dep. Variables:
    - Total trial to criterion
    - Total short-term retention score (24 hours): isolation test, sentence test, new context test, word designation test
    - Total long-term retention score (3 weeks): isolation test, sentence test, new context test, word designation test
  - Major independent variable:
    - Word-alone vs. word-sentence

- Sample:
  - 115 kindergarten children in six kindergarten classrooms selected from schools in middle class areas of a Greater Vancouver school district, British Columbia, Canada
  - 43 eliminated due to illness, transfer, non-English speaker, or recognition of one or more target words
  - 72 left, randomly put into three groups: word-alone, word-sentence, control, with 24 in each

- Material:
  - Eight target words: television, our, fixing, tooth, fell, out, Nicki's, needs.
  - Forms two sentences:
    - Nicki's tooth fell out.
    - Our television needs fixing.
- Procedure:
  - Occasion 1: trial to criterion
  - Occasion 2: short-term retention
  - Occasion 3: long-term retention

- Results:
  - Total trials to criterion: 56.5 (WA), 26.4(WS)
  - Total short term retention: 11.3(WA), 15.33 (WS)
  - Total long term retention: 10.3(WA), 15.9(WS)
- Statistical tests of significance
- Support H2.

## Additional things to do …

Read <u>Rash, Judy, Terry D. Johnson & Norman Gleadow (1984)</u> very carefully. At this point, you should be able to understand most of the issues in this research article. Try to think about theories, unit of analysis, time dimension, major variables, sampling, and mode of observation used in this study. Are there things that can be improved in ideal situations? What are the limitations of this study caused by its research design?