Threats to Internal & External Validity

Y520

Strategies for Educational Inquiry

Robert S Michael

Why Use Experimental Designs?

Man prefers to believe what he prefers to be true.

— Francis Bacon

Threats to Internal & External Validity

- Is the investigator's conclusion correct?
- Are the changes in the independent variable indeed responsible for the observed variation in the dependent variable?
- Might the variation in the dependent variable be attributable to other causes?

Why is Internal Validity Important?

- We often conduct research in order to determine cause-and-effect relationships.
- Can we conclude that changes in the independent variable caused the observed changes in the dependent variable?
- Is the evidence for such a conclusion good or poor?
- If a study shows a high degree of internal validity then we can conclude we have strong evidence of causality.
- If a study has low internal validity, then we must conclude we have little or no evidence of causality.

Variables & Internal Validity

- **Extraneous** variables are variables that **may** compete with the independent variable in explaining the outcome of a study.
- A **confounding** variable is an extraneous variable that **does** indeed influence the dependent variable.
- A confounding variable systematically varies or influences the independent variable and also influences the dependent variable.
- Researchers must always worry about extraneous variables when they make conclusions about cause and effect.

Necessary Conditions for Causality

- Three conditions that are necessary to claim that variable A causes changes in variable B:
  - Relationship condition: Variable A and variable B must be related.
  - Temporal Antecedence condition: Proper time order must be established.
  - Lack of Alternative Explanation Condition: Relationship between variable A and variable B must not be attributable to a confounding, extraneous variable.

Example:
Necessary Conditions for Causality

- A correlation (aka, relationship) exists between coffee drinking and the likelihood of having a heart attack.
- Are we justified in concluding that coffee drinking *causes* heart attacks?
- Cigarette smoking is related to both of these variables. Individuals who drink little coffee are less likely to smoke cigarettes than are people who drink a lot of coffee.
- The observed relationship between coffee drinking and heart attacks might be the result of the third variable of smoking.
- A researcher must control the effect of smoking in order to determine if this rival explanation accounts for the original relationship.

Threats to Internal & External Validity

- The controlled or experimental design enables the investigator to control for threats to internal and external validity.
- Threats to **internal** validity compromise our confidence in saying that a relationship exists between the independent and dependent variables.
- Threats to **external** validity compromise our confidence in stating whether the study's results are applicable to other groups.

Threats to Internal Validity (a)

- **History:** Did some unanticipated event occur while the experiment was in progress and did these events affect the dependent variable?
  - History is a threat for the one group design but not for the two group design.
  - In the one group pre-post test design, the effect of the treatment is the difference in the pre-test and post-test scores. This difference may be due to the treatment or to history.

Threats to Internal Validity (b)

- **History:**
  - Is not a threat for the two group (treatment/experimental and comparison/control) design because the comparison is between the treatment group and the comparison group.
  - If the history threat occurs for both groups, the difference between the two groups will not be due to the history event.

Threats to Internal Validity (c)

- **Maturation:** Were changes in the dependent variable due to normal developmental processes operating within the subject as a function of time?
  - Is a threat to for the one group design.
  - Is not a threat to the two group design, assuming that participants in both groups change ("mature")at same rate.

Examples:

Threats to Internal Validity (d)

- **History:** In a short experiment designed to investigate the effect of computer-based instruction, Ss missed some instruction because of a power failure at the school.
- **Maturation:** The performance of first graders in a learning experiment begins decreasing after 45 minutes because of fatigue.

Threats to Internal Validity (e)

- **Statistical regression:** An effect that is the result of a tendency for subjects selected on the bases of extreme scores to regress towards the mean on subsequent tests.
- When measurement of the dependent variable is not perfectly reliable, there is a tendency for extreme scores to regress or move toward the mean.
- The amount of statistical regression is inversely related to the reliability of the test.

Examples:

Threats to Internal Validity (f)

- **Statistical regression:** In an experiment involving reading instruction, subjects grouped because of poor pre-test reading scores show considerably greater gain than do the groups who scored average and high on the pre-test.

Threats to Internal Validity (g)

- **Selection:** Refers to selecting participants for the various groups in the study. Are the groups equivalent at the beginning of the study?
- If subjects were selected by random sampling and random assignment, all had equal chance of being in treatment or comparison groups, and the groups are equivalent.
- Were subjects self-selected into experimental and comparison groups? This could affect the dependent variable.
- Selection is not a threat for the one group design but it is a threat for the two group design.

Threats to Internal Validity (h)

- **Experimental Mortality:** Differential loss of participants across groups.
  - Did some participants drop out? Did this affect the results?
  - Did about the same number of participants make it through the entire study in both experimental and comparison groups?
  - Is a threat for any design with more than one group.

Threats to Internal Validity (i)

- **Testing:** Did the pre-test affect the scores on the post-test?
  - A pre-test may sensitize participant in unanticipated ways and their performance on the post-test may be due to the pre-test, not to the treatment, or, more likely, and interaction of the pre-test and treatment.
  - Is a threat to the one group design.
  - Not a threat to the two group design. Both groups are exposed to the pre-test and so the difference between groups is not due to testing.

Examples:

Threats to Internal Validity (j)

- **Selection:** The experimental group in an instructional experiment consisted of a high-ability class, while the comparison group was an average-ability class.
- **Experimental Mortality:** In a health experiment designed to determine the effect of various exercises, those subjects who find the exercise most difficult stop participating.
- **Testing:** In an experiment in which performance on a logical reasoning test is the dependent variable, a pre-test cues the subjects about the post-test.

Threats to Internal Validity (k)

- **Instrumentation:** Did any change occur during the study in the way the dependent variable was measured? (Is a threat to the one group design; not to the two group design. Why?)
- **Design contamination:** Did the comparison group know (or find out) about the experimental group? Did either group have a reason to want to make the research succeed or fail? Often, investigators must interview subjects after the experiment concludes in order to find out if design contamination occurred.

Examples:

Threats to Internal Validity (l)

- **Instrumentation:** Two examiners for an instructional experiment administered the post-test with different instructions and procedures.
- **Design contamination:** In an expectancy experiment, students in the experimental and comparison groups “compare notes” about what they were told to expect.

Threats to Internal Validity (m)

- **Compensatory rivalry.** When subjects in some treatments receive goods or services believed to be desirable and this becomes known to subjects in other groups, social competition may motivate the latter to attempt to reverse or reduce the anticipated effects of the desirable treatment levels.
- Saretsky (1972) named this the "John Henry" effect in honor of the steel driver who, upon learning that his output was being compared with that of a steam drill, worked so hard that he outperformed the drill and died of overexertion.

Threats to Internal Validity (n)

- **Resentful demoralization.** If subjects learn that their group receives less desirable goods or services, they may experience feelings of resentment and demoralization.
- Their response may be to perform at an abnormally low level, thereby increasing the magnitude of the difference between their performance and that of groups that receive the desirable goods or services.

External Validity

- External validity refers to the degree to which the results of an empirical investigation can be generalized to and across individuals, settings, and times.
- External validity can be divided into
  - Population validity
  - Ecological validity

External Validity

- Population validity:
  - How representative is the sample of the population? The more representative, the more confident we can be in generalizing from the sample to the population.
  - How widely does the finding apply? Generalizing across populations occurs when a particular research finding works across many different kinds of people, even those not represented in the sample.

External Validity

- Ecological validity is present to the degree that a result generalizes across settings. Types include:
  - Interaction effect of testing
  - Interaction effects of selection biases and experimental treatment
  - Reactive effects of experimental arrangements
  - Multiple-treatment interference
  - Experimenter effects

Threats to External Validity (a)

- **Interaction effect of testing:** Pre-testing interacts with the experimental treatment and causes some effect such that the results will not generalize to an untested population.
- **Interaction effects of selection biases and the experimental treatment:** An effect of some selection factor of intact groups interacting with the experimental treatment that would not be the case if the groups were randomly selected.

Examples:

Threats to External Validity (b)

- **Interaction effect of testing:** In a physical performance experiment, the pre-test clues the subjects to respond in a certain way to the experimental treatment that would not be the case if there were no pre-test.
- **Interaction effects of selection biases and the experimental treatment:** The results of an experiment in which teaching method is the experimental treatment, used with a class of low achievers, do not generalize to heterogeneous ability students.

Threats to External Validity (c)

- **Reactive effects of experimental arrangements:** An effect that is due simply to the fact that subjects know that they are participating in an experiment and experiencing the novelty of it — the Hawthorne effect.
- **Multiple-treatment interference:** When the same subjects receive two or more treatments as in a repeated measures design, there may be a carry-over effect between treatments such the the results cannot be generalized to single treatments.

Examples:

Threats to External Validity (c)

- **Reactive effects of experimental arrangements:** An experiment in remedial reading instruction has an effect that does not occur when the remedial reading program, which is the experimental treatment, is implemented in the regular program.
- **Multiple-treatment interference:** In a drug experiment the same animals are administered four different drug doses in some sequence. The effects of the second through fourth doses cannot be separated from the possible delayed effects of preceding doses.

Control of Threats to Internal Validity:

In the following tables, a plus sign (+) means the threat to validity is controlled.

A minus sign (–) indicates a definite weakness in this design concerning the threat to validity.

A question mark (?) means a possible threat exits.

A blank space indicates the threat is likely not relevant.

Control of Threats to Internal Validity:
Pre-exp. Designs for Description

	Case Study	1 Grp Pre-Post	Static Group Comparison
History	–	–	+
Maturation	–	–	+
Testing		–	+
Instrumentation		–	+
Regression		?	+
Selection	–	+	–
Mortality	–	+	–
Contamination		–	–

Control of Threats to External Validity:
Pre-exp. Designs for Description

	Case Study	1 Grp Pre-Post	Static Group Comparison
Testing and X		–	
Selection and X	–	–	–
Reactive effect.		?	
Multiple X inference			

Control of Threats to Internal Validity: Experimental Designs

	Post-test Only, Ctrl Group	Pre-Post Test, Ctrl Group	Solomon Four Group
History	+	+	+
Maturation	+	+	+
Testing	+	+	+
Instrumentation	+	+	+
Regression	+	+	+
Selection	+	+	+
Mortality	+	+	+
Contamination	+	+	+

Control of Threats to External Validity: Experimental Designs

	Post-Test Only, Ctrl Group	Pre-Post test, Ctrl Group	Solomon Four Group
Testing and X	+	–	+
Selection and X	?	?	?
Reactive effect	?	?	?
Multiple X inference			

Control of Threats to Internal Validity:
Quasi-Experimental Designs

	Time Series	Non-Equiv Ctrl Group	Separate Sample Pre-test / Post-test
History	–	+	–
Maturation	+	+	–
Testing	+	+	+
Instrumentation	?	+	?
Regression	+	?	+
Selection	+	+	+
Mortality	+	+	–
Contamination	+	–	–

Control of Threats to External Validity:
Quasi-Experimental Designs

	Time Series	Non-Equiv Ctrl Group	Separate Sample Pre-Test / Post-Test
Testing and X	–	–	+
Selection and X	?	?	+
Reactive effect	?	?	+
Multiple X inference			