The Internal Validity of Research

Internal Validity

- Addresses the extent to which the research design:
  - Ensures that the only possible explanation for the results (dependent variable) is the effect of the independent variable (manipulated)
  - Rules out confounding variables that provide alternative explanations for the observed effect
    - limit the effects of extraneous variables (keeps them from becoming confounding variables)
      or
    - estimate their effects

Confounding Variables

- Varies systematically along with manipulation of independent variable so that you can not tell which is causing the change in the dependent variable
- E.g., Exposing two groups of children to video – then observing them for aggressive play in one hour following video. One group (A) is exposed to violence in the video and the other (B) watches a neutral video.
- However, Group A was recruited from a sample of children from abusive homes. Group B was recruited from a privileged daycare with low levels of domestic violence in the homes.
- If some children had been exposed to violence in the home but were randomly distributed between the groups, this would NOT constitute a confounding variable

Types of Confounds

Treatment confounds

- The IV in an experiment is confounded with another variable
  - E.g., In exercise study, participants who ran indoors also heard music; those who ran outdoors did not
  - Participants should have been exposed to only one treatment
  - Instead they were exposed to two:
    - music + indoor location
      or
    - no music + outdoor location

Types of Confounds

Natural confounds

- Some variables are associated with certain other variables in nature
  - People who share demographic characteristics undergo a common set of experiences
  - Leads to a set of common attitudes, values, and other characteristics
    - E.g., "Millennial" college students share a set of core personality traits: Sheltered, Confident, Team-Oriented, Conventional, Pressured, Achieving

Types of Confounds

- Measurement Confounds
  - Occur when a measure assesses more than one hypothetical construct
    - E.g., Measure of Depression also assesses anxiety
  - Shows lack of discriminant validity for measure
  - Which construct is really correlated with your outcome?

Time-Related Threats to Validity

History: The possibility that events external to the research have affected the behavior being studied

- Example: A researcher wants to assess whether making campus smoke-free decreases smoking
- Possible history effect: Cost of cigarettes increases 25 percent at same time smoke-free designation introduced

Time-Related Threats to Validity

Maturation: A natural change over time that could serve as an alternative explanation for changes in participants' responses

- Researcher studies whether computer-based training program improves children's spelling
  - Over time period of study, children's cognitive development makes learning spelling easier

Time-Related Threats to Validity

Testing: Taking the pretest affects scores on the posttest independently of the effect of the intervention

- Researcher studies whether completing prejudice-reduction program decreases prejudice
  - Participants are motivated to show decreased prejudice and try to "do better" on posttest
- Practice
- Fatigue

Time-Related Threats to Validity

Instrumentation change: The measure used to assess the dependent variable changes over time, leading to artificial differences in scores at different points in time

- Researchers code infants' nonverbal reactions to stimuli
  - Coders experience "observer drift"
- Heart rate monitor becomes less sensitive to detecting changes in HR

Time-Related Threats to Validity

Statistical regression (regression toward the mean): An increase or decrease of initial extreme scores from pretest to posttest

- When scores are measured a second time, they will, on average, be less extreme
- Researcher studies whether low GRE scores can be raised by training program
- Extremely low scores likely to improve (without treatment) at retest due to regression toward the mean

Control Groups in Pretest-Posttest Research

Control groups can be used to assess pretest-posttest change

- Experimental group represents the effect of experimental treatment over and above the time-related threats to internal validity
- Without control group, changes due to IV cannot be identified

Pretesting

- Pretesting can
  - verify pre-experimental equivalence of experimental and control group
    - If not equal – can measure relative difference in change
  - allow comparison of drop-outs with those who complete study

Selection Threats to Validity

Selection bias: Occurs when research participants in control group differ from those in experimental group (or participants in two experimental groups differ for reasons other than manipulation of the IV)

Due to: Nonrandom assignment, such as when people volunteer for different groups

Forms of Selection Bias

Preexisting natural groups assigned to condition

- Example: All residents at one Alzheimer's treatment center assigned to experimental group; all residents of a different center assigned to control group

- Differences between the groups aren't random
  - One center may be urban, the other rural
  - One may accept Medicaid recipients, the other doesn't

Forms of Selection Bias

Mortality: People drop out of study as it is conducted

- Differential mortality: Members of one condition of experiment more likely to drop out
  - Example: In test of experimental drug with severe side effects, more participants in experimental group drop out
- Pretesting can help assess the effect of mortality
  - Demographic characteristics of dropouts can be compared to completers

Threats to Internal Validity

- Diffusion of treatment
- Compensatory behavior
  - Rivalry (John Henry effect)
  - Equalization
- Resentful Demoralization

Reactivity

Participants' scores on measures result from their reaction to the situation rather than from the effects of the independent variable

- E.g., children do poorly on cognitive task b/c they are distracted/excited by the novelty of testing situation

Can affect both internal and external validity

Sources of Reactivity

Novelty effects: Any aspect of a research situation that is new or novel and can induce reactivity

- "New" environment, such as laboratory, can affect behavior
  - E.g., Being attached to physiological equipment may cause apprehension

Novelty effects can be reduced by:

- giving participants time to adjust to situation or equipment
- observing behavior in a natural setting long enough for participants to stop editing their behavior
- ethical use of deception

Sources of Reactivity

Evaluation apprehension: The anxiety people feel when they believe someone is judging their behavior

- High levels motivate people to avoid adverse judgments
  - Can do so by behaving in ways that lead to positive judgments
  - Especially likely in presence of authority, like researchers
  - May distract them from task

Evaluation Apprehension can be reduced by:

- not labeling experimenters as psychologists
- not describing study as one that assesses ability or personality
- avoiding giving verbal feedback to participants' responses (e.g., "good job," "not quite right")
- putting participants at ease
- demonstrating familiarity with participants' culture

Artifact

- Something other than the independent variable that
  - affects the dependent variable
  - provides an alternative explanation for the results
  - E.g., During a study of effects of reputation (selfish versus generous) potential donors on chimpanzee begging behavior, some of the donors were recognized by the chimpanzees as vet techs

Demand Characteristics

Information present in the research situation that allows participants to form their own hypotheses about the purpose of the study

- Assumptions about study's purpose can affect responses
  - Can operate even if assumptions are incorrect

Participant Roles

In response to demand characteristics, participants may take on one of these participant roles

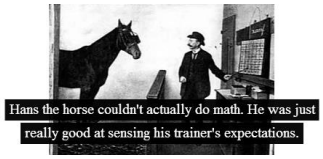
- *The Good Participant*: "Helping" the researcher by responding in the "right" way
- *The Negative Participant*: Psychological reactance or other factors lead participants to show their "independence"
- *The Apathetic Participant*: Participants are unmotivated, so pay little attention to research task or their responses
- *Anxious Participant*
- Participants may enact "*faithful*" role
  - Following instructions and behaving as they usually would in situation

Controlling Demand Characteristics

- Reduce cues that might produce artifactual results
  - Avoid obvious experimental manipulations
  - Use between-subjects designs when possible
- Use pilot studies and/or post-experimental interviews to explore whether demand characteristics were operating
- Use ethical methods of deception if necessary and applicable
- Use methods that motivate research participants
  - Avoid those that induce psychological reactance
  - Remind participants that their participation is voluntary
  - Inform participants about importance of study

Experimenter Expectancy Effects

Occur when an experimenter's beliefs about how participants *should* perform on the research task affects how participants *actually* perform on the task



Experimenter Expectancy Effects

Participants' responses can be affected if experimenters unconsciously

- introduce variations into research procedures that result in experimental and control groups being treated differently
- indicate verbally or nonverbally which answers are "right" or "wrong"

Observer Expectancy Effects

- In observational studies, expectancy can influence how participants' behaviors are recorded
  - Researchers who expect one group to perform better or worse might observe this, even if performance did not differ by group
  - Example: Rosenthal & Fode's (1963) "maze bright"/"maze dull" rat studies

Rosenthal & Fode

- How do you think the experimenter effect worked in this case? If the experimenters' expectations really did lie behind the differences in performance in the two groups how might these expectations have been communicated to the rats?
- 2. Why were the participants led to believe that the point of the whole exercise was to give them practice in handling rats and in duplicating experimental findings?

Experimenter Expectancy cont.

- Harvard Test of Inflected Acquisition
  - Pygmalion effect
  - Rosenthal and Jacobson
  - Children placed in lists of "high" and "low" achievers randomly
  - Children shown as "ready to bloom" showed greater gains in IQ test at end of year

Ways to Control Expectancy Effects

- Ask experimenters to follow a detailed script when interacting with participants
- Monitor experimenters' behavior to ensure it does not change over time
- Ensure the experimenter does not know which condition participant is in
- Avoid data snooping
- Minimize the experimenters' role, when possible

External Validity

Addresses the question of whether the results of a particular study hold up under new conditions

- Also referred to as generalizability

External Validity

Generalizing across (Generalizability): Do the results of a study pertain equally to more than one setting, population, or subpopulation?

- Of interest to basic researchers
- Focus is on general principles of behaviors
- If results don't generalize as expected, sets boundary conditions of theory

External Validity

Generalizing to (ecological validity): Do the results of a study pertain to a particular setting or population?

- Of interest to applied researchers
- Do the research conditions mimic those found in the natural setting of interest?
  - Strictly speaking, only results of research conducted in setting itself are considered valid for application to that setting

Components of External Validity

Component	*Focus*	*Issue Addressed*
Structural	Methodology	Are the results consistent across settings, procedures, populations, etc.? (both G and EV)
Functional	Psychological Processes	Are the psychological processes that operate in research settings similar to those that operate in natural settings? (EV)
Conceptual	Conceptual	Is the research question important in the applicable natural setting? (EV)

More on Structural Component of External Validity (Generalizability):

Researchers want to generalize findings across

- Time
  - Do the relationships between variables change as a function of historical events or social change?
- Cultures
  - Concerns the shared history, traditions, and worldview of a social group

Setting Factors

- Physical settings vary from study to study
  - Affects participants' responses to IV
  - Thus, affects generalizability of results
- If environmental factors interact with IV, responses can vary from setting to setting
  - .e.g., outdoor runners experience fluctuations in temperature that indoor runners do not

Setting Factors

Researchers' characteristics can affect participants' responses, including

- gender
- race/ethnicity
- personality
- values
- Experience

- Can also be influenced by attributes of co-participants

Setting Factors

Reactivity due to experimenter characteristics controlled by use of

- multiple experimenters
- computer-based instructions/interviewing
- detailed experimental script

Participant Sample Factors

Are participants different from the general population?

- Likely if participants are chosen from a
  - Restricted sample
    - Only one category of persons studied (e.g., men, Christians)
  - Convenience sample
    - Usually college students
      - Western
      - Educated
      - Industrialized Countries
      - Rich
      - Democratic

Research Volunteers Tend to Be

- Higher in
  - education levels
  - SES
  - IQ scores
- Less
  - authoritarian
  - conforming
- From smaller towns
- Younger
- More
  - sociable
  - excitement-seeking
  - unconventional
  - interested in religion
  - altruistic
  - self-disclosing
  - maladjusted
- Female

Person by Situation Interactions

- Different types of people might respond differently to same IV
  - If characteristics of volunteers interact with IV, generalizability to non-volunteers is poor
  - Even those required to participate choose which experiments to participate in
    - Important to keep track of dropout rates
- If so, results may apply only to the people represented in the sample

Research Procedure Factors

- Responses can vary from operational definition to operational definition
- Limits generalizability

Possible solutions

- Use multiple operational definitions of construct

More on Ecological Validity

Variations in lab study procedures affect generalizability across studies

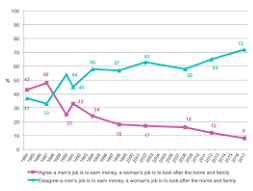
- Most laboratory studies
  - are artificial
  - isolate people from their accustomed environments
  - present them with relatively simple and time-limited tasks
  - involve interactions among strangers
  - are limited because there are usually no consequences for poor performance or harm done to others

More on Ecological Validity

Factors that increase experimental control also

- may require artificial means of communication between participants
- can limit the applicability of results
- can result in situations that would never happen in everyday life
  - Example: In everyday life, product preferences are not expressed by responding to 5-point rating scales
  - Example: In everyday life, people deal with multiple sources of input simultaneously; in the lab, sources are usually limited

Time and External Validity



- Some behaviors are cyclic
  - Frequency rises and falls at regular intervals
    - Example: Water skiing accidents occur more in summer; snow skiing accidents occur more in winter
- Some IVs require time to have an effect
  - Example: Programs to reduce bullying in schools may take weeks or months to show effect
- Some behaviors change over time
  - Are affected by social norms, history, and cultural changes

More on the Functional Component of External Validity

To what extent does the behavior elicited in the research situation mirror the behavior elicited in natural situations?

- Example: Using simulator to train for defensive driving may not translate into behavior on the road

- Is sometimes difficult to gain access to behavior in natural settings
  - Example: Researchers seldom have access to people's interactions with their physician

Assessing Generalizability

Multiple tests of a hypothesis can be used to
- determine the circumstances under which a hypothesis is supported or not
- identify the boundary conditions of the hypothesis
- generate and test new hypotheses

Assessing Generalizability

Replicating the effect of an IV enhances the construct validity of the principle being tested
- Replications should be done using different
  - operational definitions
  - settings
  - populations
  - procedures

In Defense of Laboratory Research

- Whether research has ecological validity is an empirical question
  - Whether results generalize is testable
  - Results of meta-analyses generally support the generalizability of lab research
- Research conducted in a natural setting also might not be generalizable

Are These Arguments Convincing to Critics of Laboratory Research?

Probably not
- Debate has been going on for 50 years
- Positions held are a function of personal beliefs about the scientific process
- These viewpoints relate to researchers' goals and interests
  - Basic researchers are more interested in establishing boundary conditions
  - Applied researchers are more interested in specific settings

The Experimenter's Dilemma

The Trade-off Between Internal and External Validity

- The more tightly controlled an experiment, the stronger its internal validity (ability to infer cause and effect). However, tight experimental control makes the experiment more unique and less like other settings, thereby lowering external validity.
- Experimenters almost always opt for internal over external validity.