Internal Validity, External Validity, Pitfalls

What You Should Learn

- Define the concept of “confounding”
- Explain how confounds threaten the internal validity of research, and recognize confounds in summaries of research.
- Define the “Campbell and Stanley” threats to internal validity.
- Explain the role of control groups in protecting internal validity.
- Explain various research design techniques to protect internal validity.

What you should learn (Cont.)

- Define the concept of “reactivity” as it applies to research settings, and describe the sources of reactivity.
- Explain measures that can be taken to control reactivity.
- Understand and control “demand characteristics.”
- Explain the effects of participant roles on research.

What you should learn (Cont.)

- Understand investigator effects that can lead to invalid research conclusions
- Explain Research Assistant effects that can lead to invalid research conclusions
- Understand concept of External Validity

Internal Validity: Definition

- Internal validity refers to the extent to which we can accurately state that the independent variable produced the observed effect.
- If
  - effect on dependant variable only due to variation in the independent variable(s)
- then
  - internal validity achieved

Example of the Issue

- Investigating effects of tutoring on grades
- Compare those who receive tutoring with those who do not receive tutoring
- Tutored students do better
  - brighter
  - receive more nonspecific attention
  - don’t stay out late
- Internal validity is questionable

Confounding

- Extraneous variable
  - any variable other than IV that influences DV
- Confounding
  - occurs when an extraneous variable systematically varies with variation in IV
  - the extraneous variable affects the DV
  - plausible alternative explanation
  - tutoring and intelligence vs. birth order

Controlling Extraneous Variables

- Can eliminate some extraneous variables
- Most must be controlled
- Example: CVC and learning method
  - control for word association
- Difficulty lies in identifying the variables

Variables We Know That Must Be Controlled

- History
- Maturation
- Testing
- Instrumentation
- Statistical Regression
- Selection
- Mortality
- See Cook and Campbell (1979) for others

History

- An extraneous variable occurring between pre- and post-measurement of the DV
- Refers to specific events, other than IV
- Example: Attitude-change study
  - measure attitude toward gun control
  - attitude change manipulations
  - Shooting occurs at two office buildings, 27 people die
  - measure attitude toward gun control

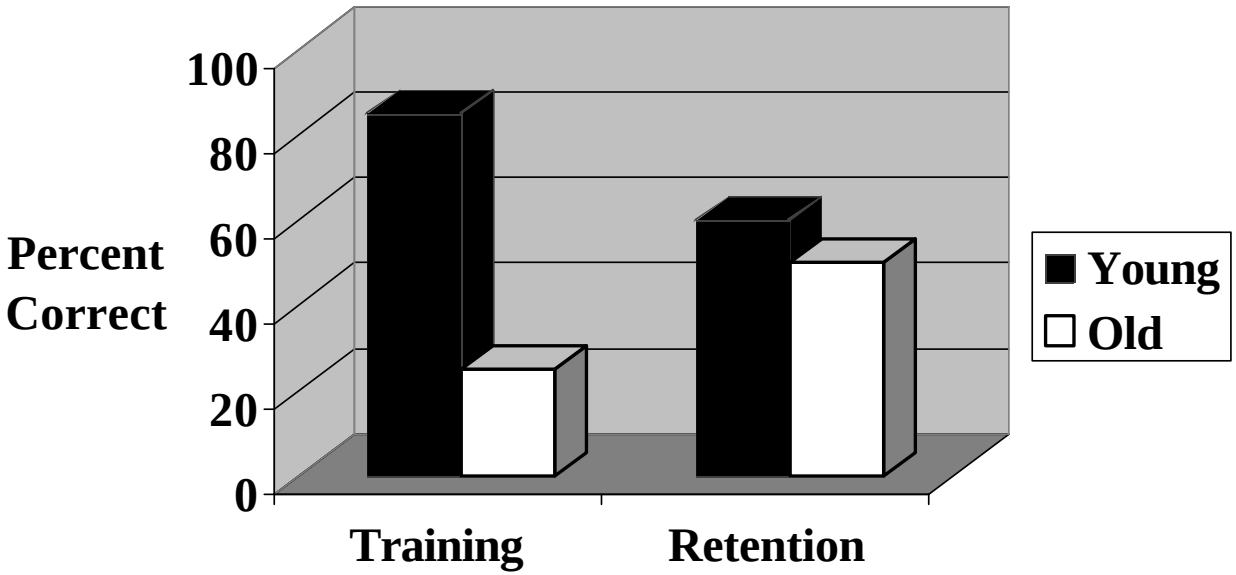
History: Another Example

- Dietary change on violence in institutionalized juveniles
- New group of inmates
- Record behaviors for three months
- Change diet
- Record behaviors for three months
- Violence declines after diet change

Maturation

- Changes in biologial and psychological conditions that occur with passage of time
- Refers to the internal changes of individual that occur due to pasage of time
- Consider: Retention of learning and effects of age on retention
  - First assess performance after 6 continuous hours of practice
  - Test performance one month later
- What was “discovered”

Maturation Example



Maturation Again

- Testing benefits of Head Start Program
- Pretest to establish “ability” of slow learners
- Set up special room to motivate these kids
- One year later retested same kids
- Found 1.75 years improvement for the 1.0 year in the program.
- Fame and fortune awaited the researchers…..

Testing

- Repeated measurement on the same variable leads to improved performance because of
  - learning
  - practice
    - general learning
    - specific learning
  - conjecture about the research
- What are examples?

Instrumentation

- Changes that occur due to changes in the assessment of the DV
- Does not refer to participant changes
- Refers to the changes that occur during process of measurement
- Changes in researcher
  - – becoming more skilled, or tired
- Changes in the instrument itself

Statistical Regression

- The lowering of extreme high scores and the raising of extreme low scores
- Change scores problematic for many reasons, this is one
- Does not mean people “regress toward mediocrity” but the statistical effect of regression toward mean can cause interpretation problems

Illustration of Statistical Regression Effect

Participant	Pretest	Selected Participant	Pretest	Posttest
S1	110	S1	110	103
S2	46	S3	123	116
S3	123	S8	105	98
S4	92			
S5	59			
S6	73			
S7	99			
S8	105			
S9	67	S2	46	57
S10	84	S5	59	63
S11	61	S9	67	70
S12	96	S11	61	65

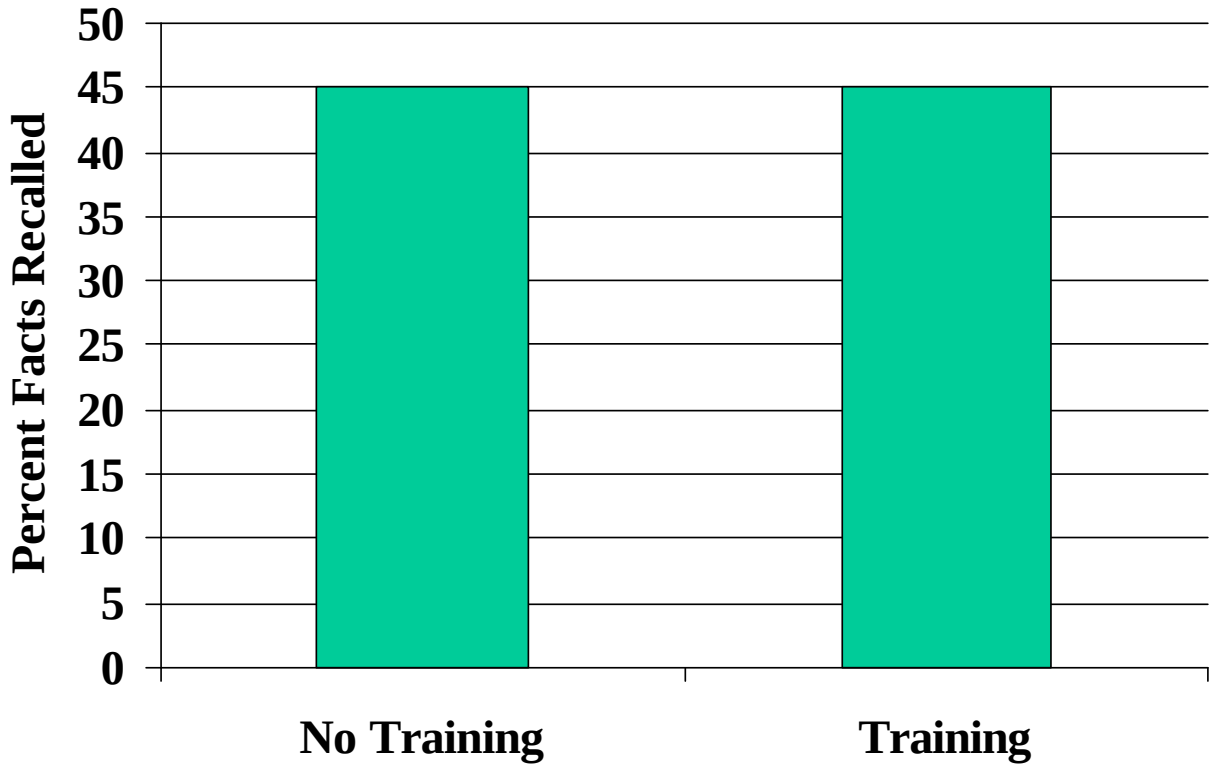
How Can Regression to Mean Lead to Interpretation Problems?

- Score high on first exam, score less well, on average, on final exam
- Score low of first exam, score better, on average, on final exam
- Interpretation:
  - The instructor brings everyone to average
  - The instructor can only teach gifted students
  - And so on…..

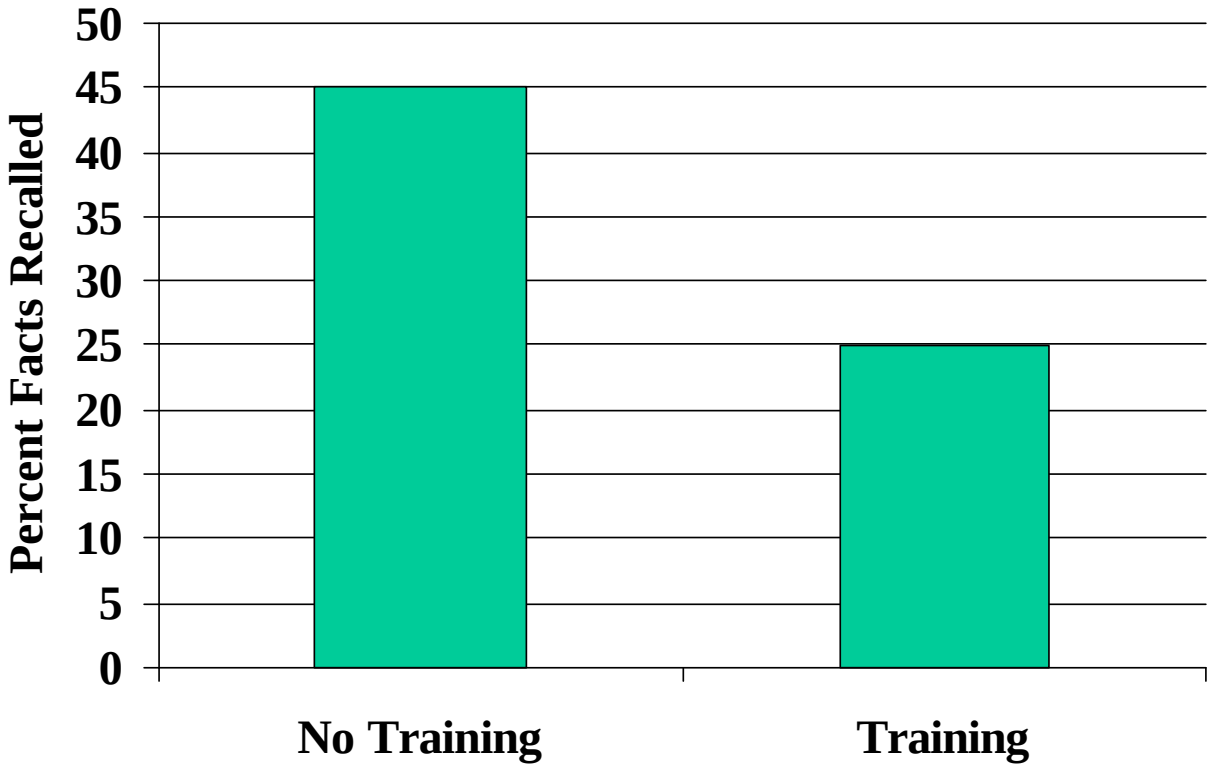
Selection

- The choice of participants for the various treatment groups is made on the basis of different criteria
- Ideally sample is randomly chosen from a population then randomly assigned to treatment groups
- If not, rival hypotheses are introduced
- Example:
  - “Morning” group and “Evening” group

Selection


Percent Facts Recalled
50
45
40
35
30
25
20
15
10
5
0
No Training
Training

Selection


Percent Facts Recalled
50
45
40
35
30
25
20
15
10
5
0
No Training
Training

Mortality

- A differential loss of participants from the various treatment groups in the study
- Problem is not just loss but differential loss such that differences may be due to who is left not treatment
- Examples:
  - training method and retention
  - Longitudinal studies and effects of age
    - **“only the strong survive”?**

Conclusion: Threats to Internal Validity

- The threats we covered are not exhaustive
- Internal validity may be threatened from multiple sources
- Your job as scientist:
  - ensure alternative explanations can be ruled out
- Checklist approach not really possible
- You must think

External Validity

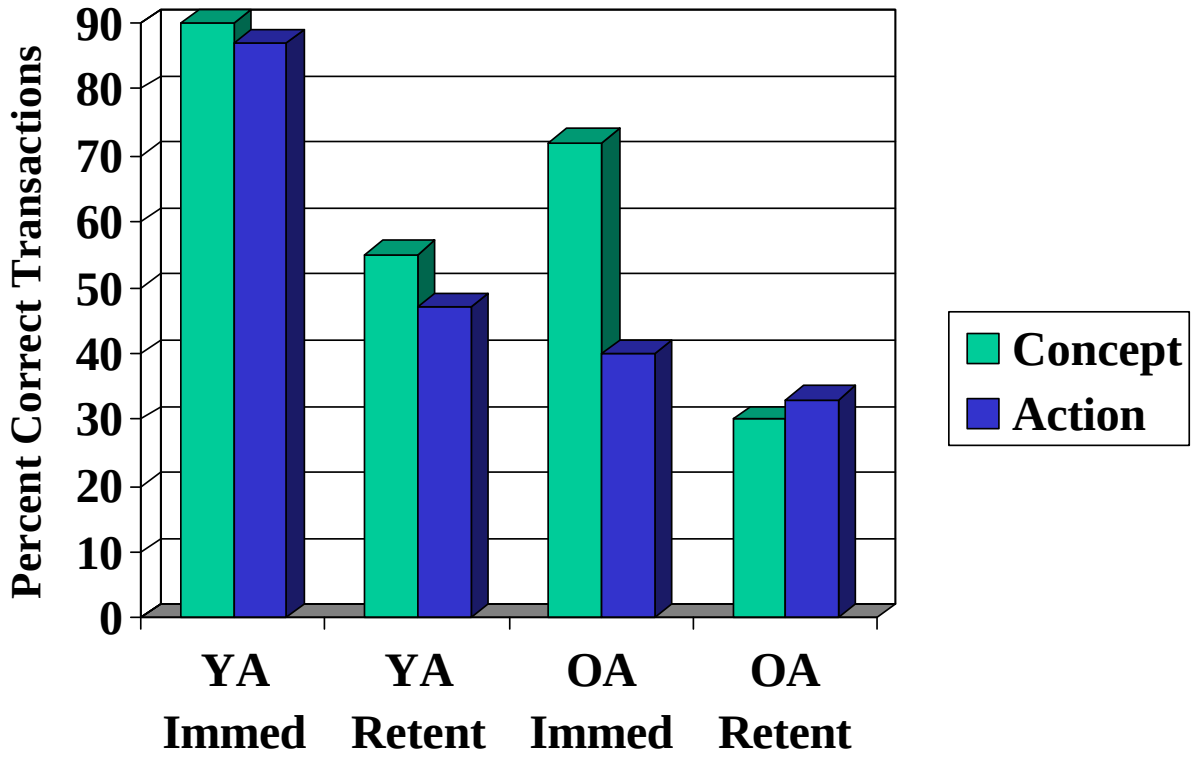
- What is external validity?
- Relates to generalizing your findings
  - **to** or **across** target populations
  - **to** or **across** tasks
  - **to** or **across** environments
- Campbell and Stanely: “ the ability to generalize to or across exemplars of a particular to the entire class of a particular”

List of Some Threats to External Validity

- This list not exhaustive
- This list not meant to serve as a checklist
- This list should stimulate your thinking when you are concerned with generalizations
  - of your own work
  - of the work of others

Examples of Threats

- Treatment-Attribute Interaction
- Treatment-Setting Interaction
- Multiple-Treatment Interference
- Pretest Sensitization
- Post-test Sensitization



From Mead & Fisk, 1999 Age Related Training Study

Examples of Threats

- Treatment-Attribute Interaction
- Treatment-Setting Interaction
- Multiple-Treatment Interference
- Pretest Sensitization
- Post-test Sensitization

Summary Internal/External Validity

- What is Internal Validity?
- Internal Validity Threats
  - History
  - Maturation
  - Testing
  - Instrumentation
  - Statistical Regression
  - Selection
  - Mortality
- What is External Validity?
- External Validity Threats
  - Treatment-Attribute Interaction
  - Treatment-Setting Interaction
  - Multiple-Treatment Interference
  - Pretest Sensitization
  - Post-test Sensitization

Artifacts and Pitfalls

- Still concerned with Internal Validity
- Focus now on issues emanating from different aspect of research process:
  - the Participant
  - the Research Assistant
  - the Principal Investigator
- Why focus on what might go wrong?

Participant Effects

- Perfect participant exists in our dreams
- Participants come to study with
  - expectations, biases, personalities, etc.
- Type of participants
  - the good, the faithful, the negativistic, the apprehensive

Participant Effects: What to Do

- Be aware that these various kinds of participants exist.
- Give no cues that lead to a particular kind of behavior
- If you find an ***exaggerated*** type of participant
  - keep notes in your study log book
  - you may wish to exclude prior to looking at that participant's data

Research Assistant and Investigator Effects

- Discuss 10 pivotal points
- Two main questions
  - At what point in the research process can study go astray giving misleading results
  - What steps can be taken to avoid pitfalls

Define Investigator and Research Assistant

- Can be same person, usually not
- Investigator
  - decides study is to be conducted
  - how it is designed and carried out
  - how data analyzed and interpreted
- Research Assistant
  - conducts study
  - tests participants
  - records, enters data

The Major Pitfalls: Investigator Effects

- Investigator Paradigm Effect
- Investigator Research-Study Design Effect
- Investigator Loose Procedure Effect
- Investigator Data Analysis Effect
- Investigator Fudging Effect

The Major Pitfalls: Research Assistant Effects

- RA Personal Attributes Effects
- RA Failure to Follow Procedure Effects
- RA Incorrect Recording Effect
- RA Fudging Effect
- RA Unintentional Expectancy Effect

Investigator Paradigm Effect

- What is a paradigm and why important
- When do problems arise
  - results inharmonious with accepted paradigm view as not acceptable
  - Example
- Recommendations
  - be aware of assumption
  - be aware of pitfall of “proving” theory
  - thoroughly test multiple alternative hypotheses
    - “studying hypotheses”
    - Not “substantiating theories”

Investigator Research-Study Design Effect

- Same paradigm, similar theory
- Different results because of design
- Examples
  - complexity of design
  - within vs. between
- Recommendations
  - need to place emphasis on fact that results are dependant on way study is designed

Investigator Loose Procedure Effect

- Degree of imprecision of study protocol
- Recommendations
  - provide precise specifications as to how study is to be conducted
  - plan for contingencies to ensure everyone treated same
  - standardize things like
    - how to greet each participant
    - what to do if participant interrupts procedure

Investigator Data Analysis Effect

- Investigator has control of and responsibility for data analysis
- Seven types of data analysis problems
  - No preplanning
  - Failing to report non-supporting data
  - Inappropriate post-mortem analyses
  - Not correcting for multiple analyses
  - Selective reporting of significant results
  - Not reporting failures to replicate
  - Checking only non-confirming analyses

Investigator Data Analysis Effect

- Recommendations for improvement
  - If not planned comparisons report all data
  - Do not change alpha level in “mid-analysis”
  - Substantiate post-mortem tests by further research
  - Avoid “probability pyramiding”
  - Plan study with manageable number of IV/DV

Investigator Fudging Effect

- For sake of completeness we will discuss
- Occurs when reported results are not actual results
- Not just outright faking but also
  - "pushing the data"
  - selectively discarding
  - changing a p value from .07 to .05
  - selectively trimming data
- Even if person just suspected, treated as pariah

RA Personal Attributes Effect

- Attribute of research assistant (e.g., gender or ethnicity) can affect participants' responses on specific study task.
- But complex effects for
  - whether attribute of RA affects responses on wide variety of task
  - whether multiple attributes add or interact
- Recommendation
  - realize effect is real, design for internal and external validity

RA Failure to Follow Procedure Effect

- If RA deviates meaningfully from established procedure then the published study is misleading. It is not the study that was actually conducted.
- RA can vary in way they conduct study
- Within an RA they may test different participants differently
- Recommendation
  - design for internal and external validity

RA Incorrect Recording Effect

- Failing to correctly record participants' responses
  - random error or systematic error
- Where:
  - recording answers to ability tests given one on one
  - recording events during usability testing
  - scoring and entering data
- Why
  - not careful
  - desire to "meet expectations"

RA Fudging Effect

- Not too difficult to document
- When most likely to happen
  - "hired-hand" RAs
  - "piece rate" workers
  - not engaged in research effort

RA Unintentional Expectancy Effect

- Do expectations and desires lead to unconscious, unintentional effects?
- Perhaps in ways such as tone of voice, posture, facial expressions, etc.
- But most studies fail to show this effect if other factors controlled
- When interpretation required and criteria ambiguous, problem can arise

General Review of Section