

Unit Two: The Basics of Empirical Research

- Research Design
- Measurement

Group Design Questions



- **Experimental conditions**
 - Does cognitive behavioral therapy work better than nondirective counseling?
 - Are 12 group sessions as effective as 16 sessions?
 - How will change in the eligibility for free lunch effect the number of people applying for the program?
 - Will reduced case loads improve the recidivism rate for families with child abuse?

Types of Research



■ Exploratory

- Little is known about the area.
- Discover which variables are relevant.

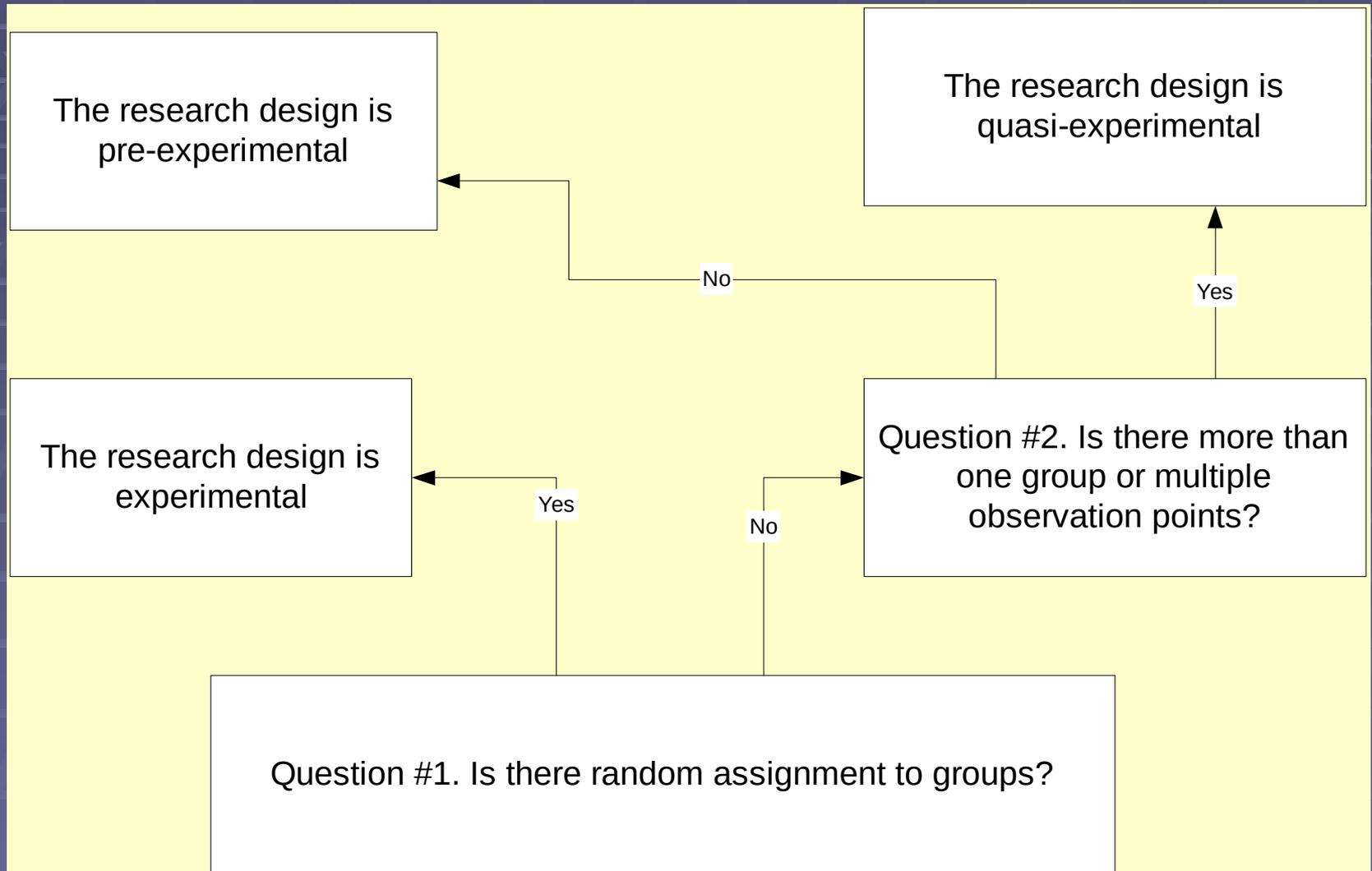
■ Descriptive

- Describing a phenomenon in a holistic way.
- How many are involved; what characteristics
- What processes occur; how often.

■ Explanatory

- Focus on “why” things happen; prediction.
- Explain the relationship between variables.
- Test theory.

Types of Group Designs



Start

Design Notation



- X – exposure to the independent variable
- O – Observation (or measurement)
 - O_1 – First observation or measurement
- E – Experimental group
- C – Control group
- R – Random assignment (Not matching!)
 - All elements have an equal chance of selection for the experimental and control group.
- NR – Not random assignment
 - Nonequivalent group or Comparison group

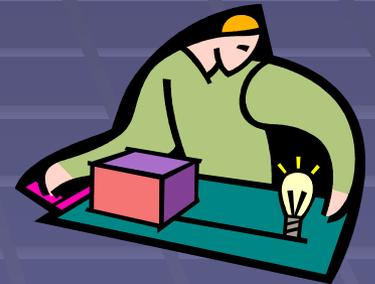
Pre-experimental Design

- **Cross sectional (“one shot”) case study**
 - Notation: O
 - Exploratory or descriptive studies.
 - One measurement of one group.
- **One group, post-test only**
 - Notation: X O₁
 - One measurement following an independent variable (intervention).
 - Research Question: Has a minimum standard of achievement been met?



Quasi-experimental Designs

- Provides comparison measures
 - Additional measures over time or additional groups but without random assignment.
 - Cannot rule out alternative explanations for change in the dependent variable.
- One group pre- / post-test
 - Notation: 0×0
- Post test only with nonequivalent groups
 - Notation: $X \ 0$
 0



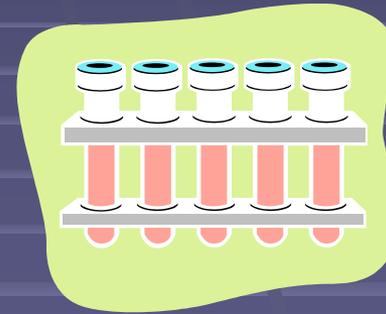
Quasi-experimental Designs (con't.)

- Pre-post test with comparison group

- Notation: $O \quad X \quad O$
 $0 \quad 0$

- Multigroup post-test only

- Notation: $X_1 \quad O_1$
 $X_2 \quad O_1$
 $X_3 \quad O_1$



- Simple Time Series

- Multiple measurements of the dependent variable before and after the introduction of the independent variable.
- Notation: $O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_5 \quad O_6 \quad O_7 \quad O_8$

Quasi-experimental Designs (con't.)

- **Multiple Time Series Design**

- Notation: $O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8$
 $O_1 O_2 O_3 O_4 \quad O_5 O_6 O_7 O_8$

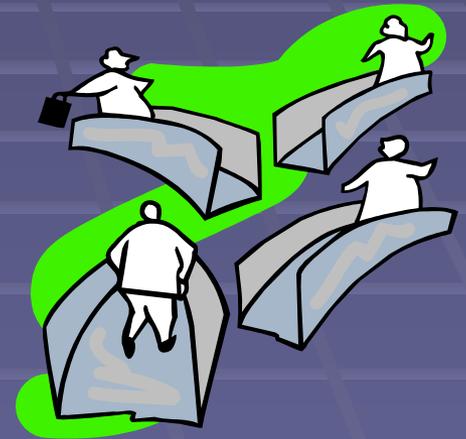
- **Uninterrupted Time Series Design**

- Observational research without independent variable.
- Notation: $O_1 O_2 O_3 O_4 O_5 O_6$



Experimental Designs

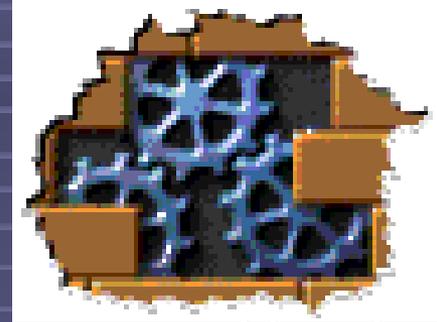
- *A controlled method of observation in which one or more independent variables is changed in order to assess its causal effect on one or more dependent variables.*
 - Random assignment to groups controls rival hypotheses.
- **The classic experiment**
 - E: R O₁ X O₂
C: R O₁ O₂
 - Experimental and Control group.
 - Random assignment.
 - Independent variable given to experimental group.



Experimental Design (con't.)

■ Post-test only design

- Notation: $R \quad X \quad O_1$
 $\quad \quad R \quad O_1$



- Random assignment assumes equality at pretest.
- Use if testing or instrumentation is an “issue.”

■ Solomon Four-Group Design

- $R \quad O_1 \quad X_1 \quad O_2$
- $R \quad O_1 \quad \quad O_2$
- $R \quad \quad X_1 \quad O_1$
- $R \quad \quad \quad O_1$

- Used to test the effects of pre-testing

Experimental Design (con't.)

- **Limitations of experimental designs**
 - High need for control limits generalization to “real life” – decreased external validity.
 - High level of attrition may compromise group equivalence.
 - Threat to internal validity makes conclusions suspect.



Threats to Internal Validity

- **History** – extraneous events
 - Events taking place between pretest and post test influence the results.
- **Maturation or the Passage of Time**
 - A cut finger heals with time.
 - People become older, wiser, more tired with or without the independent variable.



Internal Validity (con't.)



■ Testing

- Learning to take the test
 - SAT test results get better the second time.
- Learning because of the test.
 - E.g., reading about parenting techniques after a parenting test.

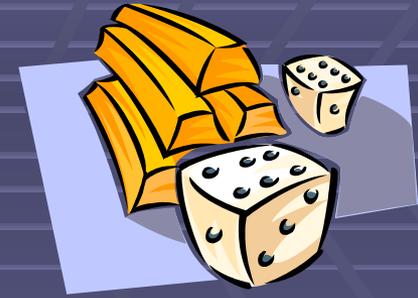
■ Instrumentation

- Are pre-test and post-test the same?
- Does taking the test change responses?
- Raters get bored or better at rating with time.

Internal Validity (con't.)

■ **Statistical regression**

- Extreme scores move toward the mean
 - Genius mother has less smart daughter
 - Very tall parents have shorter kids
 - Very dissatisfied becomes less dissatisfied.



■ **Selection bias**

- Are groups comparable?
 - Day versus evening clients
 - “Stacking the deck”: all troublemakers in control group.

■ **Experimental mortality**

- Are drop-outs different from those who remain?
 - How many dropped out?
 - Who's left?

Internal Validity (con't.)

- **Treatment Contamination**
 - Control group learns from experimental group
 - Kids teach each other on the playground.
 - Therapists learn techniques from each other.
- **Resentful Contamination**
 - Control resents being in the control condition
 - Causes greater differences between groups.



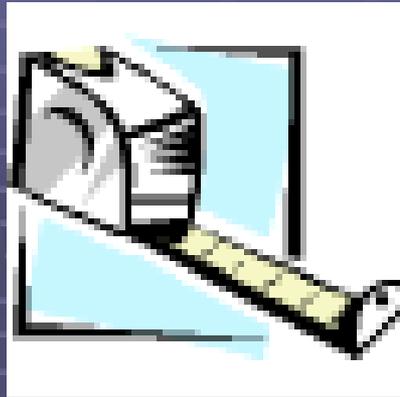
External Validity

- The extent to which studies are reflective of real life.
 - Experiments are planned; life isn't.
 - Experiments are simple; life isn't.
 - Generalization must take into account:
 - Age, race, sex, class, education, reading level
 - Organizational setting, geography, culture
 - TO WHOM DO RESULTS APPLY?

Threats to External Validity

- **Reactivity (Hawthorne Effect)**
 - Being in an experiment or being observed changes “typical” behavior.
- **Researcher bias**
 - Researchers see what they want to see
 - Researchers evaluate experimental and control groups differently.
 - Use double-blind method when possible.
- **Multiple treatment interference**
 - Effects of two or more treatments cannot be separated.
 - Those in treatment are also talking with minister or coach.
 - Parent education group may also be reading about parenting.





Measurement

Measurement

- ***The process of assigning numbers or other symbols to characteristics or attributes of a concept according to specified rules.***
- **Tools**
 - Ruler, thermometer, stopwatch, weight scale.
- **Scales**
 - Beck Depression Scale
 - Family Environment Scale
 - Self-esteem Scale
 - Conflict Tactics Scale
- **Self report / Other report**
- **Direct observation**
- **Analysis of interview or other records**



Variables

- **Something that can vary**
 - If we study self-esteem, it varies from low levels to high levels.
 - Attitudes vary from strongly disagree to strongly agree.
- **Measurement choices**
 - Ask questions (Primary data)
 - Observe (Primary data)
 - Consult existing records (Secondary data)
- **What is Measured**
 - Frequency (how often)
 - Duration (how long)
 - Magnitude (how much)



Levels of Measurement

- **Nominal** – categories
 - Sex, race, religious affiliation, political party
 - No numerical meaning.
- **Ordinal** – more or less: groupings
 - Satisfaction or Agreement: high, medium, low.
 - Education: high school, some college, college.
 - Quantitative meaning is unclear.
- **Interval** – category distance has meaning
 - I.Q. score. (But no zero I.Q.)
- **Ratio** – equal intervals and true zero.
 - Age, number of children, number of days absent.



Which Level of Measurement?

■ *Nominal, Ordinal, Interval, Ratio ?*

- Democrat or Republican
- Grade point average
- Exercise: None, A little, Moderate, Great deal
- SAT score
- Marital Satisfaction Scale from 15 to 150
- Which research class: 1=M, 2=W, 3=S
- Classmate Popularity: rank order from most popular to least popular
- Income (actual)
- Income: 1 = <\$20,000; 2=\$20,000 – \$50,000; 3=>\$50,000



Survey Research



- A systematic way of gathering data using a structured format.
 - Measure what respondents know, believe, feel, or how they say they behave
- **Self-administered questionnaire**
 - A survey that is mailed to potential respondents or that is given in person
 - Web-based surveys are increasingly used
- **Interview Schedule:**
 - survey is administered face to face or by telephone.
 - CATI (Computer Assisted Telephone Interview)

Types of Surveys



- **Structured – maintain uniformity**
 - Closed-ended questions with forced choices
 - “Are you a Democrat, Republican, or Other?”
- **Semi-structured**
 - Interviewer is free to clarify the questions and follow-up on the participant’s responses, but questions are pre-arranged.
 - Use of open-ended questions – no choices
 - “How has caring for your elder parent affected you?”
- **Unstructured**
 - Totally open-ended interview with no predetermined direction for questions.
 - “Tell me a little about yourself.”

Observational Research

"Actions speak louder than words"

- **Pure observation**

- Researcher remains apart from the observed:
 - Avoids **social desirability** – tendency to say the socially acceptable thing or to tell people what you think they want to hear.

- **Participant observation**

- Observer becomes an active participant in the research setting.
 - **Reactivity** – people tend to change behavior when they know they are being observed.
 - **Observer Inference** – Observer interprets events from her/his own point of view.

Existing Records

- Paper and electronic documents, databases, and other non-document artifacts of human behavior. E.g., census data, agency files, police records.
 - Advantages:
 - Inexpensive, fast
 - Disadvantages:
 - No control over the types or accuracy of the data
 - Different data sets may be difficult to compare or combine



Apples or Oranges



■ Quantitative

- Consistent; standardized
- Large samples; generalizable results
- Reductionistic; limited number of predefined variables

■ Qualitative

- Open-ended, in-depth, insight-oriented
- No preconceived answers to questions
- Seeks patterns and processes
- No limit on types of data or questions

Both provide useful information

Standardized Instruments



- **A tool to measure a concept**
 - **Item:** single indicator of a variable, e.g., one question.
 - **Index:** simple addition or mean of items.
 - **Scale:** includes different weighting of items.
- **May be *normed***
 - Information about how various large groups scored on the instrument.
 - Need to assess the generalizability of the norms and when they were established.

All instruments have some degree of error.

Measurement Error



- **Random error**

- Errors that are not consistent or patterned.
- Anything that can't be controlled but is likely to be equally distributed among subjects.
 - Health, motivation, random answers, typos
 - Feelings about the researcher or subject matter
 - Reading level
- Thought to balance out across the study.

Systematic Error

- *Patterned or consistent error*
 - *Bias in how questions are asked*

Interviewer smiles when we get the right answer.

 - Keep data collectors and subjects “blind.”
 - *Response set*

Tending to agree or disagree.
 - *Social desirability*

Saying the nice or “right” thing to look good.
 - *Question wording*

“Assistance to poor” versus “Welfare.”

Systematic Error (cont'd.)



- *Question order or questionnaire introduction*
 - Question: “Which are the most prevalent social problems today?”
 - Answers: CRIME, poverty, health
- *Subject selection*
 - Reading level doesn't match subject ability
- *Cultural / language bias*
e.g., I.Q. tests, SAT

Always pretest your instrument

Validity

- *Does the instrument measure what you think it's measuring? Real meaning.*
- **Face validity**
 - Does it makes sense?
 - Will people take it seriously?
- **Content Validity**
 - How are the items selected?
 - Do they represent the full range of the concept?



Validity (cont'd.)



■ Criterion Validity

- Same results as something known to measure the concept (e.g., skin-based blood sugar).
- *Concurrent Validity*: test against another valid measure.
- *Predictive Validity*: ability to predict future events, e.g., suicide scale predicts attempts.
- *Discriminant Validity*: ability to discriminate known groups, e.g., abusers from non-abusers, depressed from not depressed.

Validity (cont'd.)



- **Construct Validity**
 - Correlation with expected concepts and propositions, NOT with others outside of the theory.
 - Abuse correlates with stress and social isolation but not with hair color or IQ.
 - Technology skills correlate with years of computer ownership, but not with birth order or life satisfaction.

Reliability



- **Stability:** Measures don't change:
 - Thermometer, bathroom scale
- **Equivalence:** all items measure the same concept.
 - Spelling test shouldn't have math items.
- **Internal consistency**
 - Items testing the same concept should have similar scores.
- **Reliability does not equal accuracy**
 - Bathroom scale off by 15 pounds is reliable.
 - Thermometer is broken, but is reliably wrong.

Promoting Reliability



- **Questions**
 - Keep it simple
 - Be sure people know the answer or opinion
- **Interviewers / Raters**
 - Training, regular check-ups
- **Testing Reliability**
 - Test-retest
 - Parallel forms or Split half
 - Interrater reliability (two or more)
 - Intrarater reliability (test same rater twice)
 - Internal consistency (consistency of item scores with each other)

Relationship

- If reliable, not always valid.
- If not reliable, never valid.
- If not valid, may be reliable.
- If valid, always reliable.



Finding Measures

- Program objectives and research needs determine measures to use.
- Places to find measures:
 - Measure is part of journal article.
 - Published in book of measures.
 - Contact the test author and request permission to use their test.
 - Use a directory published by scientific and professional associations like APA.
 - If these attempts to locate the author fail, contact the publisher holding the copyright to the original material and request permission from the publisher.
 - Search the Internet.

Designing Measures



- **Maximize response rate**
 - Use a professional looking instrument
 - Avoid the cluttered look and small type.
 - Be explicit in explaining to respondents why the research is being conducted.
 - Make assurances that the results will be kept confidential, and that no services will be withheld or denied even if the respondent chooses not to participate.
 - If potentially sensitive items are to be asked, keep them to a minimum and position them later in the instrument.

Design (cont'd.)

- Avoid asking questions that the target population is not expected to know anything about.
- Avoid biased questions
 - e.g., “Do you believe that all forms of abortion, the killing of pre-born babies, should be illegal?”
- Keep open ended questions to a minimum.
- Keep survey as short as possible.
- Begin with interesting questions first.
- Explain transition to next section.



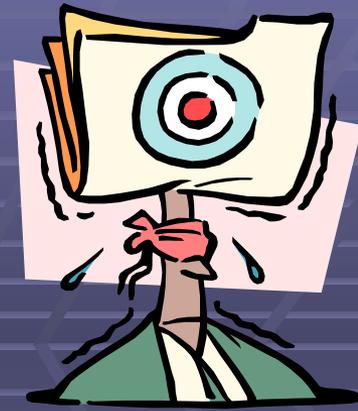
Design (cont'd.)



■ Minimize Error

- Consider the audience: age, education, etc
- Have simple, clear questions
 - Keep questions short, under 20 words.
 - Avoid slang terms.
- Position response categories vertically underneath the question stem rather than side-by-side.
- Use the highest level of measurement possible.
 - Ask exact age rather than a category.
- Don't ask for unavailable information
 - Asking college seniors, "How much did you study in the 7th grade?"

Design (cont'd.)



- Avoid double questions
 - *Do you support abortion or are you a conservative?*
- Avoid negative wording
 - *Do you not think the government should pay for school lunches?*
- Use clear instructions.
- Have instruments translated when necessary.
- Reduce social desirability
 - Anonymous
 - Remind users that there are no right answers

The Politics of Measures

- **Definition of measures is political**
 - Define “Marriage.” Who is included and who is not? How is it measured?
- **Measures may not be generalizable to all groups**
 - Many measures are tested on white, middle-class college students.
 - Some medicines are tested only on men, but their dosages may be different for women.
- **Measurement may be used to reinforce stereotypes**
 - Girls can’t do math.
 - Men have higher SAT scores. (Is the test biased?)

