Experimental Design

Part I
Richard S. Balkin, Ph. D, LPC-S, NCC

Overview
• Experimental design is the blueprint for quantitative research and serves as the foundation of what makes quantitative research valid.
• Too often, consumers of research may focus on the rationale for a study and the discussion of the results. However, without proper methods, a quantitative study has no meaning.

Foundations of Quantitative Research
• Instead of merely focusing on the statistical test, the end result so to speak, a quantitative study is based on a strong foundation of a generalizable sample, reliable and valid instrumentation, and appropriate research design.
Consider the following scenario

A school counselor wishes to improve upon a program for students with low academic achievement. Currently, the school counselor leads a psychoeducational group on study skills. Based upon the responses from the participants, the school counselor believes that students with academic problems are having difficulty in areas outside of study skills. The school counselor decides to implement a group counseling program in addition to the study skills program and wishes to investigate whether group counseling would have a greater effect in increasing academic achievement among the participating students.

Consider the following scenario

A researcher needs to consider how a program or intervention is deemed effective. For example, if the participants obtaining group counseling and study skills interventions are performing better, can the school counselor conclude it is due to the program? Progress could be due to either intervention, the combined intervention, or another variable not measured. The mere passage of time could be a reason for improvement.

Defining Variables

In an experimental study there are two types of variables:

- Independent variable (I will abbreviate this as the IV)
- Dependent variable (I will abbreviate this as the DV)
Independent variable

- The independent variable in classical experimental design is the variable that is being manipulated.
- In the previous scenario, participants could be randomly assigned to (a) study skills only, (b) group counseling and study skills, and (c) no intervention.
- Participation in group counseling and study skills, study skills only, and no intervention are the various levels of the independent variable.

Dependent variable

- The dependent variable is the variable that is being measured
- In the previous scenario, participants could be measured on change in grade point average (gpa).

Experimental Validity

- Experimental validity refers to the process in which results are generalizable because the factors that have been tested or manipulated (the independent variable) truly effect a change in an outcome (dependent variable), and the results of the study can be applied to settings outside of the experimental setting (Best & Kahn, 2006).
- In order to determine the effectiveness of a program or intervention, appropriate methods for sampling (i.e. random sampling), measuring the outcome, and establishing the validity of the findings are necessary.
For example, Kelly, Halford, and Young (2000) studied the effects of a short-term intervention related to alcohol abuse. The researchers assigned 16 participants to a treatment group and 16 participants to begin the intervention after one month (control group). After the initial month, the researchers reported that the treatment group had made statistically significant progress in reducing alcohol and depressive symptoms. The control group did not show any progress with respect to alcohol abuse but did show a statistically significant reduction in depressive symptoms.

Can the researchers conclude that the intervention was more successful in treating depressive symptoms?

The researchers cannot conclude that the intervention was more successful in treating depressive symptoms. Depressive symptoms appeared to diminish whether or not an intervention occurred. In order to determine whether a program or intervention is effective, experimental conditions are required. These conditions may include the implementation of treatment and control groups, random assignment, and measures to insure experimental validity.

Random assignment refers to the equal likelihood that a participant will be assigned to a treatment, control, or comparison group.

Random assignment is similar to random sampling. In random sampling, participants from an accessible population had an equal chance of being selected for a study.

In random assignment, each participant has an equal chance of being selected for a particular intervention or no intervention at all.
Random assignment

- In the group counseling/study skills example, the school counselor would want to randomly assign participants to group counseling and study skills and study skills only. The school counselor may opt to also include a control group (no services) but this should be handled carefully as there are ethical issues in denying or delaying services.

Why random assignment is important

- Random assignment helps ensure experimental validity by providing a measure in equalizing groups.
- Random assignment protects against selection bias; no group is predisposed to a treatment or intervention.
- Random assignment also protects the influence of confounding variables, variables and attributes that are not being measured but may influence the results of a study (Gall et al., 2006).

Why random assignment is important

- By using random assignment, a confounding variable is theoretically dispersed equally across all groups.
- “Random assignment is meant to make control and experimental groups equivalent” (Vogt, 2007, p. 96).
Three Types of Design

- There are three major types of experimental design:
  - (a) pre-experimental in which there is no random assignment and no comparison group
  - (b) quasi-experimental in which a comparison group is used but no random assignment
  - (c) true experimental in which both random assignment and a comparison group is utilized.

Pre-experimental designs

- Pre-experimental designs are easy to implement but the findings may not be reflective of what is being measured because of the absence of a comparison group.
- The meaningfulness of a study utilizing a pre-experimental design should always be questioned as the conclusions cannot be generalized.

Pre-experimental designs

- In a pre-experimental design, participants would be placed in both study skills and group counseling. The school counselor would evaluate whether or not academic performance improved as evidenced by changes in grade point average, standardized test scores, etc.
- If academic progress was demonstrated by the participants, could the school counselor conclude the intervention was successful?
Pre-experimental designs

- No.
- Without any comparison group, the researcher does not know if group counseling, study skills, or simply the passage of time results in change of the DV (such as gpa).

Quasi-experimental designs

- Quasi-experimental designs are very common in social science research because of the use of non-manipulated factors, which prevent random assignment of the independent variable.
- Special attention to ensuring group equivalence is pertinent.

Quasi-experimental designs

- In a quasi-experimental design, participants would be assigned to study skills and group counseling and study skills only, but the assignment would not be random.
- For example, the school counselor could use intact groups based on their class schedule. Before the school counselor evaluates whether or not academic performance improved as evidenced by changes in grade point average, standardized test scores, etc., statistical tests should be run to make sure that neither groups has significantly higher or lower test scores or grade point average, as this could bias the results.
Quasi-experimental designs: Group Equivalence

- Group equivalency refers to each group being similar prior to a manipulation of an IV
- For example, if one group of students with an average GPA of 1.65 was assigned to a study skills group and another group of students with an average GPA of 2.50 was assigned to study skills and group counseling would we have equivalent groups?

Quasi-experimental designs: Group Equivalence

- No.
- And if the study skills and group counseling participants had greater increase in GPA, we would not know if it was due to the intervention or the fact that the participants were already performing better.
- When group equivalency is demonstrated, results can be generalized as the effect of the independent variable can be evaluated.

True experimental designs

- True experimental designs are more complicated to implement but the results are generalized more easily
  - Random assignment promotes group equivalency
  - The effect of the independent variable can be ascertained more easily due to the presence of a comparison group.
True experimental designs

- In a true experimental design, participants would be randomly assigned to study skills and group counseling and study skills only. Thus, all participants would have an equal chance at being selected for study skills and group counseling or study skills only.
- The school counselor would evaluate whether or not academic performance improved as evidenced by changes in grade point average, standardized test scores, etc.

Threats to experimental validity

- Before examining models of experimental design, it is important to understand why these models exist—to substantiate experimental validity.

Threats to experimental validity

- There are two types of experimental validity.
  - *Internal validity* is the extent to which the independent variable(s) truly effect the change in the dependent variable.
  - *External validity* is the extent to which the study can be generalized to other settings and populations (Best & Kahn, 2006).
Threats Internal Experimental Validity

- Threats to internal validity are due to extraneous variables
- Extraneous variables are any variables that are not controlled for that can affect the outcome.
- In social sciences, it is impossible to control for every conceivable extraneous variable.
- However, true experimental designs can minimize the effects of extraneous variables.

To understand the effect of an extraneous variable, consider the group counseling example at the beginning of the lecture. Even if the groups were randomly assigned and participants in study skills and group counseling had greater change in academic performance than participants receiving study skills only, could the school counselor be certain that the change in academic performance was due to the added group counseling intervention?

Campbell and Stanley (1966) identified nine factors that could threaten internal validity. Most of the threats listed in this section can be controlled for through random assignment.

- Maturation
- History
- Testing
- Instrumentation
- Statistical regression
- Selection bias
- Interaction of selection and maturation
- Mortality
- Experimenter bias

Review pages 172-175 in your textbook. You will be responsible for understanding these concepts. Please email me if you have questions.
Maturation

- Maturation refers to the change in the participants over time. Participants may change simply due to time passing during the study.
- Other factors that may involve maturation are changes in emotional, intellectual, and/or physical functioning or fatigue from participation in the study (LaFountain & Bartos, 2002).

---

Maturation

- Remember the alcohol abuse/depression study on slide 10?
- In the Kelly et al. (2000) study, participants in the control group exhibited a statistically significant decrease in depression despite not having any intervention.
- This would be known as a maturation effect.

---

History

- Unplanned events that occur during the study can have an effect on the outcome.
- Participants may experience an event outside of the experimental setting that influences the outcome of the experiment (LaFountain & Bartos, 2002).
For example, a researcher is doing a study on high school athletes and moral development. During the study, the football team throws a party to celebrate a great victory. The school principal receives a call the next day that several football players were intoxicated and a female student was sexually assaulted. Such an event was beyond the control of the researcher. However, if the event had the same effect on both treatment and control groups, the effect of the event may be equalized. Otherwise, results are confounded due to the unplanned event (Best & Kahn, 2006).

This type of threat is common in designs that utilize a pretest. The actual pretest may affect future performance on a posttest. Participants may be more knowledgeable about how to answer so that progress appears more evident.

For example, using practice tests to prepare for a standardized exam may improve scores because the participant is more familiar with the question format, as opposed to being more knowledgeable about the material. However, if both the experimental and control groups are affected similarly, then the effect of the pretest is controlled. When the same test is administered repeatedly, or a series of different tests, instruments, or observations are utilized, then the testing effect is sometimes referred to as a sequencing effect. The sequence of the administrations may affect the validity of the experiment. The sequence of the treatments might be responsible for the change in the dependent variable rather than the treatments themselves.
In the group counseling example, consider the implications if the measure of academic performance was change in grade point average. Poor academic performance for students in remedial courses may be quite different from poor academic performance from students in more accelerated coursework. Thus, grade point average may not be a good measure of change in academic performance because each group member has a different set of courses. Problems with instrumentation may also develop when rating systems are used. Raters may have different standards or be influenced by other variables unrelated to the study.

For example, Riniolo, Johnson, Sherman, & Misso (2006) examined the influence of physical attractiveness on student evaluations for professors and found that instructors perceived as physically attractive by students had stronger evaluations than professor who were not perceived as physically attractive. Variations in student perceptions of physical attractiveness affected professor ratings, despite the fact that raters are not asked to consider physical attractiveness in the evaluation of an instructor.
Statistical regression
- Statistical regression refers to baseline (very low scores) and ceiling effects (very high scores).
- This problem is often seen in studies where a repeated measure is used.
- For example, in the group counseling example, students with very low academic performance will likely score higher simply because repeated measures over time tend to move toward the mean.
- Students who score very low scores are more likely to score higher on follow-up evaluations.

Consumers of research should utilize results cautiously when the participants are selected based on very low or high scores (Best & Kahn, 2006).
- The key to preventing, while never totally eliminating, statistical regression is to have variability within the sample (LaFountain & Bartos, 2002).
- Remember that baseline and ceiling effects occur when scores are at an extreme.
- If participants tend to score at one extreme or the other, groups should be divided and separate analyses may need to be conducted with participants at either extreme.

Selection bias
- Selection bias is a common problem in pre-experimental and quasi-experimental designs in which random assignment does not occur or intact groups are utilized.
- In such cases the experimental and control groups may not start at the same level or with similar scores or characteristics.
- (See slides on Group Equivalence: 22-23)
Selection bias

Consider what might happen in the group counseling example if the participants receiving study skills only had an average grade point of 1.8 and participants receiving study skills and group counseling had an average grade point of 1.2. The group with the higher grade point average may fair better because they already have more skills. Even in cases where random assignment is used there is no guarantee that the groups are equal.

Researchers can assess the influence of selection bias in a study by utilizing pretests to ensure group equivalence at the beginning of a study.

Researchers can also employ matching groups, which entails using a pretest and matching participants with equivalent scores. Participants are randomly assigned to separate groups in order to assure equivalent groups (LaFountain & Bartos, 2002).

Interaction of selection and maturation

This type of threat may occur due to different attributes between the various groups.

For example, if students in the study skills group were mostly in third grade and students in the study skills and group counseling group were mostly in fifth grade, the results may be erroneous if the age of the students is not controlled.

Interaction of selection and maturation

Interaction with selection is a common problem in social science research.

Demographic variables such as sex, ethnicity, and socioeconomic status are often studied, but such studies may lack random assignment because you cannot randomly assign participants to a variables such as sex—it is already predetermined.

In other words, simply because group differences exist does not mean that the differences are due to belonging to a specific group.
Mortality

- Participants in research studies may not complete a study.
- As a matter of fact, it is unethical to force or influence participants to complete the study if they wish to withdraw.
- If dropouts are occurring in one group more than another group, the researcher may need to consider whether the groups are still equivalent.

- If a large number of students in the group counseling and study skills group dropped out but the study skills only group remained intact, the equivalency of the groups may have been compromised.
- Attrition in long-term studies is expected. When attrition occurs in longitudinal research, comparisons with the initial sample may be compromised, especially if sample size is small. The remaining participants may not be representative of the original sample (LaFountain & Bartos, 2002).

Experimenter bias

- Experimenter bias occurs when the experimenter predisposes participants to a particular treatment.
- Experimenter bias is likely to occur when random assignment is not utilized.
Experimenter bias

- If a school counselor, for instance, believed in the value of group counseling and did not use random assignment, the school counselor may select participants that he/she believes are good candidates for group counseling, thereby affecting the outcome of the study.

Experimenter bias

- Experimenter bias is difficult to eliminate in social science research, especially when participants are selected to receive an intervention based on need.
- It is difficult to ascertain the effectiveness of an intervention, such as individual counseling, when participants who receive counseling likely volunteered for it.
- Naturally, the intervention is likely to have an impact because the participant wants the intervention to work.

Threats to external validity

- Threats to external validity are related to the artificiality of the experimental condition.
- Campbell and Stanley (1966) identified five factors that could threaten external validity.
Threats to external validity

- Interference of prior treatment
- Artificial experimental setting
- Interaction effect of testing and treatment
- Interaction of selection and treatment
- Interaction of treatment implementation

Review pages 175-176 in your textbook. You will be responsible for understanding these concepts. Please email me if you have questions.

Balkin, R. S., 2008

Interference of prior treatment

Participants who have a prior history with the treatment condition could affect the outcome of the study, particularly if such a history is not equally dispersed throughout the treatment and control groups.

Balkin, R. S., 2008

Consider this scenario:

Clients who are hospitalized in a residential setting may make considerable therapeutic progress to the extent that the client appears ready to discharge. However, upon leaving the residential program, client may regress back to their previous high risk behavior. Despite making progress in the residential setting, the therapeutic progress may not have translated well to the realities of the real world and were more a result of being in a structured setting. Thus, the therapeutic progress did not generalize to other settings.

Balkin, R. S., 2008
Interference of prior treatment

- Using the previous above, clients who have a history of receiving counseling services and are placed in an institutional setting may have an understanding of what they need to say in order to appear healthier.
- The client therefore is discharged, but the progress was very superficial.

Artificial experimental setting

- Change may occur as a result of the setting of the research.

Artificial experimental setting

- For example, participants who receive study skills and group counseling may express desire to make changes in study habits because such disclosure is positively reinforced by the group. However, upon leaving group and the counseling setting, the participant may find it difficult to change study habits due to being outside of the experimental setting where support may be much less.
Interaction effect of testing and treatment

- This is similar to the testing effect in threats to internal validity, in which the practice of taking a pre-test affects later administrations.
- In an interaction of testing and treatment, the pretest may affect the treatment and control groups differently.

Interaction effect of testing and treatment

- For example, a researcher wants to study the effects of yoga on memory. The experimental group receives a memory test, two weeks of yoga, and then another memory test. The control group receives a memory test, no treatment, and then another memory test after three weeks. Even if the treatment group scored higher on the second memory test, it is important to note that the higher memory scores could be due to the participants in the treatment group attending more to the method of the pretest, rather than the higher scores resulting from yoga practice.
- Random assignment is a preventative measure for this threat.

Interaction of selection and treatment

- Research in the social sciences is complex. Random sampling and random assignment is difficult to obtain. Social scientists depend upon cooperation from external groups, such as schools, agencies, hospitals, etc. Thus, participants are often obtained by utilizing intact groups.
- Sampling is often convenient in social science research
- For that reason, selection and treatment can compromise experimental validity. The researcher needs to be able to demonstrate that the participants in the study are truly representative of the target population.
Interaction of selection and treatment

- Reporting demographic characteristics informs readers about the generalizability of the study to other populations. When studies are conducted on populations with specific characteristics, then the study is generalizable to individuals with those characteristics.

Interaction of treatment implementation

- When developing interventions across groups, it may be difficult for a single researcher to deliver the same intervention in the same manner.

- For example, a school counselor may wish to use a colleague to provide study skills psychoeducation to one group, while the school counselor provides study skills psychoeducation and group counseling to another group.
- Even if the study skills are taught using the same materials, the presentation of the information may be qualitatively different.
- Researchers need to have procedures in place to verify that a treatment or intervention was conducted properly (Best & Kahn, 2006).