Chapter 2
Internal and External Validity

Research Methods

William D. Crano, Marilynn B. Brewer, and Andrew Lac

Created by Christopher S. Lamb to accompany Crano, Brewer, & Lac: Principles and Methods of Social Research, 3rd Edition, 2015, Routledge/Taylor & Francis

Introduction

- A wide range of potential research designs and methods will be covered in the chapters that follow
- They can be categorized roughly as involving either ***experimental*** or ***nonexperimental*** strategies:
  - Experimental methods include studies in which behavioral choices are limited or in some way constrained by the controlled manipulation of variables selected by the researcher, and characterized by random assignment of participants to different study conditions
    - **Quasi-experiment** methods do not make use of random assignment
  - In nonexperimental methods, the researcher does not systematically manipulate variables in the study, nor assign respondents randomly to different conditions of the study

Introduction

- The advantages and limitations of experimental and nonexperimental research strategies tend to be complementary
  - Effective programs of research should make use of both strategies in different stages of the research process

- Experimental studies are usually undertaken in the laboratory
  - A laboratory provides the conditions allowing the researcher to carefully regulate the context

- Nonexperimental strategies are more likely to have:
  - value of "real world" context
  - availability of mass data in developing information about human actions

Causation

- The purpose of most research studies is to investigate a hypothesized relationship between the occurrence of variation or change in one variable, A, and the occurrence of variation or change in another variable, B
- To say that there is a relationship between two such variables means that if the value or state of one variable differs or changes, we can expect that the value or state of the other will also change or differ

- For example – A: Mood Measurement; B: Weather
  - On a sunny day we observe more positive moods
  - On a cloudy day we observe more negative moods
  - *What can we suggest the relationship might be between A and B?*

Causation

- The more precise the theoretical specification of a predicted relationship, the more closely the obtained data can be matched against the prediction
- Directionality of relationships may be differentiated into three types:

  1. ***Unidirectional causation***
     - A produce subsequent changes in B, but changes in B do not influence A
  2. ***Bidirectional causation***
     - Changes in A lead to changes in B and, in addition, changing B produces changes in A
  3. ***Noncausal covariation***
     - Changes in A are indirectly accompanied by changes in B, because both A and B are determined by changes in an ***extraneous variable***, C

Causation versus Covariation

- The simple observation that when A varies, B also varies, is <u>not</u> sufficient to demonstrate which of these three possibilities is correct
- To determine that any particular causal direction exists, alternative explanations for the observed relationship must be ruled out
- When an observed relationship is found to have been mistakenly interpreted as a causal relationship of type 1 or 2, when it is actually a case of type 3, the relationship is said to be ***spurious***

Moderators and Mediators of Relationships

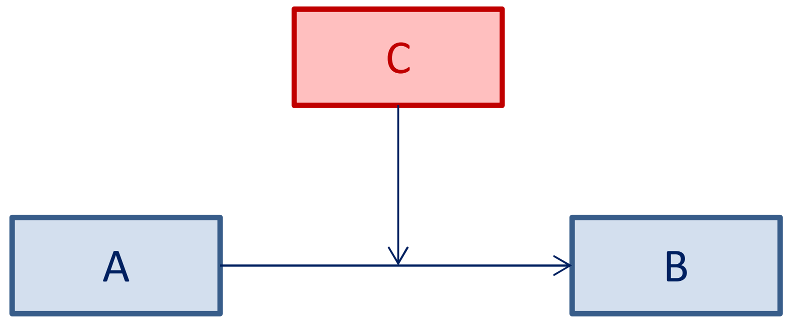
- In addition to specifying the nature and direction of a relationship under study, it also is important to distinguish between two different types of “third variables” that can influence predictive relationships—**moderators** and **mediators** (Baron & Kenny, 1986)

- Sometimes a predictive relationship can be either augmented or blocked by the presence or absence of other variables that serve as **moderators**

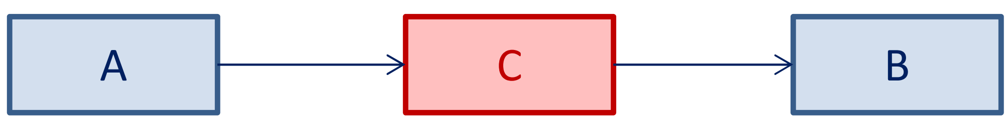
- When an additional variable (C) is necessary to complete the directional process linking two other variables (A and B) we have **mediation**

Moderators and Mediators of Relationships

- Moderation



- Mediation



Created by Christopher S. Lamb to accompany Crano, Brewer, & Lac: Principles and Methods of Social Research, 3rd

Phases of Research

- Naturalistic Observation of Human Behavior
  - Systematic observation of natural phenomena
- Classification
  - Observations are ordered according to a system of categorization, or taxonomy
- Verification Process
  - Where experimentation is possible, it is the most powerful research strategy available for determining the source and direction of relations between variables
  - The purpose of the hypothesis-testing experiment is to clarify the relationship between two (or more) variables by bringing the variation of at least one of the elements in the relationship under the control of the researcher

Independent and Dependent Variables

- In experimental research, we refer to the element that is subjected to this controlled variation as the **independent variable**
  - This variable is manipulated independently of its "natural" sources of variation
- The other element of the experimental relationship under investigation, which is observed but not controlled by the experimenter, is the **dependent variable**
  - This variable is expected to be influenced by (or is dependent upon) the manipulations of the independent variable

Independent and Dependent Variables

- The terms **independent variable** and **dependent variable** are typically reserved for experimental studies

- A more encompassing terminology, regardless of whether or not variables are manipulated, are the terms predictor variable and criterion variable (or outcome)

- The independent variable represents a specific type of a **predictor variable**, and the dependent variable is a special case of a **criterion variable** or outcome

Distinguishing Internal and External Validity

- Two forms of validity reflect upon the quality of different, but critically important, aspects of the research process:
  - **Internal validity** has to do with the certainty with which one can attribute a research outcome to a predictor
    - Internal validity is concerned with the extent to which causal inferences can legitimately be made about the nature of the relationship between predictor and outcome variables
  - **External validity** is concerned with the issue of generalizability
    - Can results can be applied to other respondent groups, to other settings, and to different ways of operationalizing the conceptual variables?
- Typically, discussion of **internal validity** is reserved for research that involves experimental methods
- Considerations of external validity or generalizability of results are equally important in evaluating the worth of experimental and nonexperimental research, and as such should be considered in both research contexts

Role of Statistics
(aka Statistical Conclusion Validity)

- The role of statistical techniques must be evaluated with respect to the general goal of eliminating or reducing the plausibility of rival alternative hypotheses for the events under investigation
  - One potential rival explanation that plagues social research at all stages of investigation is ***chance*** or random happenstance
- Statistical analyses assess the possibility that chance is a reasonable alternative explanation of any relational finding

- A study's sample is always susceptible to **sampling error**, because the outcome observed in a particular sample, by chance, is not perfectly consistent with what would be found if the entire population was used

Role of Statistics
(aka Statistical Conclusion Validity)

- Statistical inference allows us to assign a probability of the operation of chance due to sampling error as a possible explanation of any relationship discovered
- Results of a statistical inference test tell us the probability of a **Type I** error of inference
  - The likelihood that the observed effect would have been obtained if the null hypothesis was valid (i.e., there is no true relationship between the predictor and outcome variables)
- **Statistical significance** is achieved when the probability of obtaining the observed effect by chance is so low as to render the chance explanation implausible
- Tests of statistical significance provide insight into the likelihood that chance is a viable alternative explanation of our findings

Role of Statistics
(aka Statistical Conclusion Validity)

- When the probability of a Type I error is not low enough (e.g., not below $p < .05$) to rule out the null hypothesis, we have to be concerned with **Type II** errors of inference
  - Failing to reject the null hypothesis even when it is false (i.e., there really *is* a relationship between the predictor and outcome variables, but our study failed to detect it at above-chance level)
- Reducing the probability of **Type II** error requires that we design studies with sufficient power to detect an effect above and beyond random variation
- The **statistical power** of a statistical test is the probability of achieving statistical significance in detecting an effect if indeed that effect actually exists

Role of Statistics
(aka. Statistical Conclusion Validity)

Figure 2.2. Possible scenarios involving a test result and correspondence with reality.

		Actuality/Reality: Effect	Actuality/Reality: No Effect
Test/ Outcome	Effect	Correct Decision (Power)	Type I Error
	No Effect	Type II Error	Correct Decision

Created by Christopher S. Lamb to accompany Crano, Brewer, & Lac: Principles and Methods of Social Research, 3rd

Field versus Laboratory Research

- For a long time in social research, a controversy existed between the proponents of field research and those who favored laboratory studies
  - Field researchers claimed that only in real-life settings could we discover anything of value
    - Responses of participants who were studied in the cold, antiseptic environment of the laboratory could not be viewed as valid representations of the behavior being studied
  - Laboratory researchers argued that so many theoretically extraneous events occurred in the natural environment
    - So much so that one could never be certain about the true relationship that existed among any given set of variables
- As is the case with most arguments of this type, both sides were partly correct, and both were partly wrong

Field versus Laboratory Research

- There are numerous examples of laboratory studies that are so devoid of reality that their practical or scientific utility must be seriously questioned

- However, not all laboratory research is psychologically, behaviorally, and socially “unreal” (Aronson, Wilson, & Brewer, 1998)

- To argue that nothing of generalizable value can come out of the social laboratory is to deny the obvious

Field versus Laboratory Research

- There are also plenty of examples that would serve to "prove" the contentions of the critics of field research
  - The inability of these researchers to control for powerful factors that could have major influences on critical behaviors is one of the more telling problems mentioned by critics
- With the increasing sophistication evident in much contemporary field research, however, it is becoming ever more apparent that these difficulties can be surmounted
- Many recent research methodology texts have focused specifically on the complete or partial solution of the many problems encountered in field research settings.

Basic versus Applied Research

- The difference between the **basic research** and **applied research** lies in whether relatively long-term or short-term gains are expected from the outcomes of the research
  - The "applied" label refers to those research efforts that are directed toward affecting a particular phenomenon in some preconceived way
  - For basic research, the goal of each research project is to contribute to the universe of knowledge
    - In more specific terms, adding to the accumulative pattern of data that will ultimately determine the survival value of alternative theoretical interpretations of the phenomena under investigation

- The differential value of applied and basic research does not lie in any major differences in rigor of research methodology or clarity of results.

Basic Issues of Internal Validity

- The purpose of the design of experiments is oriented toward eliminating possible alternative explanations of research results (variations in scores on the **dependent variable**) that are unrelated to the effects of the treatment (**independent variable**) of interest

- When an experiment is adequately designed, changes in the **dependent variable** can be attributed to variations in the treatment, which is systematically controlled and manipulated by the investigator

- If obtained differences can be attributed directly to the experimental treatment, the study is said to have **internal validity**

Basic Issues of Internal Validity

- If factors other than the experimental treatment could plausibly account for the obtained differences, then the internal validity of the study is uncertain

- The existence of such extraneous variables is usually referred to as a **confounding** of the relationship because the effect of the independent variable on the dependent variable cannot be separated from the effects of these other variables

- A research study is said to be **confounded**, or internally invalid, when there is reason to believe that obtained differences in the dependent variable would have occurred *even if exposure to the independent variable had not been manipulated*

Basic Issues of Internal Validity

- Eight major **threats to internal validity** are:
  1. History
  2. Maturation
  3. Testing
  4. Instrumentation
  5. Statistical Regression
  6. Selection
  7. Mortality
  8. Selection-based Interactions

(see Table 2.1 for summaries)

Random Assignment and Experimental Control

- The prototype of a good experimental design is one in which groups of people who are initially equivalent (at the **pretest phase**) are **randomly assigned** to receive the **experimental treatment** or a **control condition** and then assessed again after this differential experience (**posttest phase**)

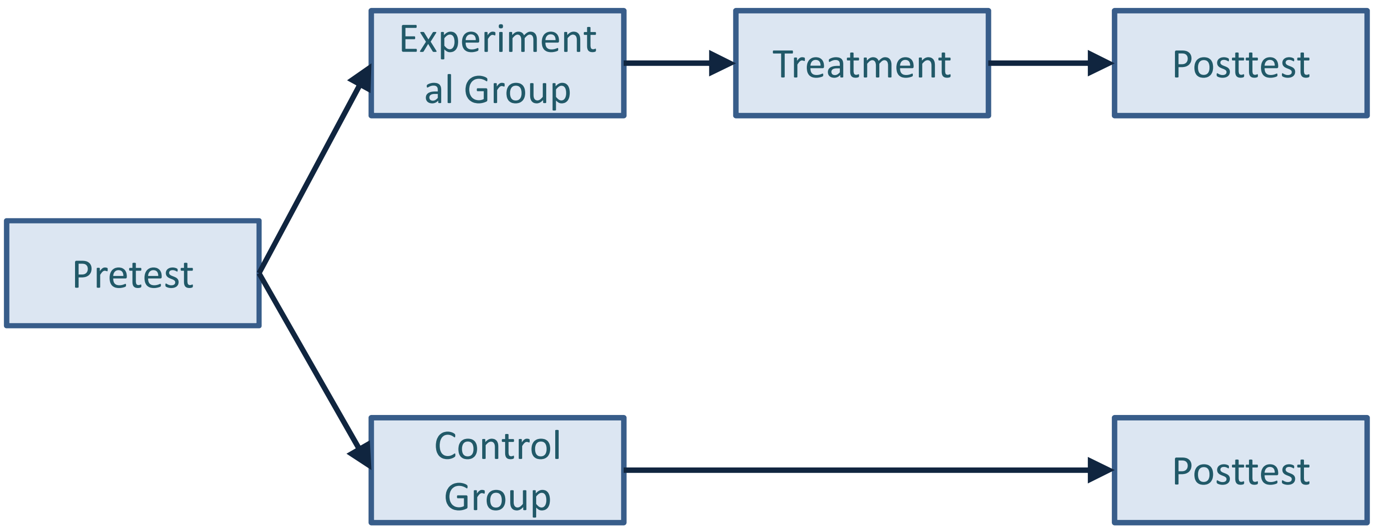


Figure 2.1. Diagram of a pretest-posttest control group design.

Random Assignment and Experimental Control

- **Random assignment** requires that all individuals available for a particular research study be able to participate potentially in either the experimental or the control group, and that only chance determines the group to which any individual is assigned
  - For example: Two groups are created as result of a coin toss. Heads assigns participant to Group A, Tails assigns to Group B

- The faith in statistical chance to place participants into groups equally is bolstered as the number of participants in a study increases

Random Assignment and Experimental Control

- When enough participants are available, chance assignment assures that there will be no pre-existing systematic differences (e.g., average age, sex, educational background, intelligence) between the groups at the onset of the study
  - There is also no reason to believe that they will experience any systematic differences during the research period, other than the treatment, whose delivery is controlled by the experimenter

- If any basis other than chance is used to assign participants to groups, then a threat to internal validity, ***selection***, may account for differences on the dependent variable

Participant Loss

- The advantages of random assignment for internal validity are assured as long as the initial equivalence of the experimental and control groups can be maintained throughout the study
  - Unfortunately, research involving human participants is never free of the possibility that at some time between random assignment to groups and the posttest, some may drop out of the experiment and become unavailable for final testing

- If there is something about the experimental treatment that enhances or otherwise affects the chances of participants dropping out of the study, a serious problem is introduced because the initially equivalent groups may become differentially selected groups by the time of final testing

Other Sources of Invalidity

- Complete control over experimental treatments is necessitated by the fact that inferences drawn from studies using a control group are acceptable only when the treatments to which the experimental and control groups are exposed differ ***only*** on the variable under consideration
  - (i.e., all other conditions are "held constant“ – keeping all possible factors or experiences similar across groups)

- A researcher's *expectations* about how the experiment will (or should) turn out can threaten an experiment even when random assignment is used
  - Such effects can be generated by:
    - *Unintentional* differences in criteria applied by observers to participants in different groups
    - *Unintentional* misinterpretation of participants' responses
    - *Unintentional* errors of data recording and analysis, and the like

Basic Issues of External Validity

- The design issues we have been discussing thus far in the chapter have been concerned almost exclusively with the **internal validity** of a research study

- For experimental as well as nonexperimental strategies, there may arise questions about the validity of interpretations of relationships obtained in any given study, particularly their applicability or generalizability beyond the results of the specific study

- These concerns constitute issues of external validity, which can be further divided into questions of
  A. Generalizability of operationalizations, and
  B. Generalizability of results to other places and participant populations

Generalizability of Operationalizations

- Concerns over the validity of operations refer to the correct identification of the nature of the predictor and outcome variables and the underlying relationship between them
  - To what extent do the operations and measures embodied in the procedures of a particular study reflect the theoretical concepts that gave rise to the study in the first place?

- Threats to this form of validity arise from errors of measurement, misspecification of research operations, and, in general, the complexity of stimulus features that constitute our variables

Generalizability of Results to other Places and Populations

- Once a research study has been completed, the investigator is usually interested in reaching conclusions that are generalizable across people and settings
  - Results applicable only to particular persons at a particular time or place are of little value to a scientific endeavor that aims at achieving general principles of human behavior

- **Generalizability** refers to the robustness of a phenomenon – the extent to which a relationship, once identified, can be expected to recur at other times, places, and populations under different environmental conditions

Generalizability of Results to other Places and Populations

- Just as the validity of measurement depends upon an adequate representation of variation in the theoretical concept, participant generalizability depends upon the extent to which the participants included in a study represent the potential variability of the human organism

- **Random selection** is required to assure generalizability of results from the sample to the entire population of relevant persons
  - The ground rule of **random selection** is that all persons in the population of interest are equally likely to be included in the research sample
  - Random selection to yield a representative sample is a rarely realized ideal in social research