VALIDITY OF QUANTITATIVE RESEARCH

Recall

> "the basic aim of science is to explain natural phenomena. Such explanations are called theories" (Kerlinger, 1986, p. 8).

Theories have varying degrees of truth.

Validity is the best approximation to the truth or falsity of propositions (Cook & Campbell, 1979).

Validity is at best approximate or tentative "since one can never know what is true. At best, one can know what has not yet been ruled out as false" (Cook & Campbell, 1979, p. 37).

And, as we have seen in examining the logic of hypothesis testing, statistical power, and the validity of outcome measures, we don't really prove that something is false. In other words, we never really prove a null hypothesis; we only fail to reject it.

"Experimental results never 'confirm' or 'prove' a theory -- rather the successful theory is tested and escapes being disconfirmed" (Campbell & Stanley, 1963, p. 35).

In other words, we fail to reject the null hypothesis.

"Varying degrees of 'confirmation' are conferred upon a theory through the **number** of *plausible rival hypotheses* available to account for the data. The fewer such plausible rival hypotheses remaining, the greater the degree of 'confirmation'" (Campbell & Stanley, 1963, p. 36).

Thus, research is a **field** of varying degrees of certainty. [Analogy: Monet's garden.]

And, "continuous, multiple experimentation is more typical of science than once-and-for-all definitive experiments" (Campbell & Stanley, 1963, p. 3). [Analogy: Brush strokes in the painting.]

Threats to validity are plausible rival hypotheses (i.e., other ways of explaining the results rather than the author(s) hypothesis).

Research design helps us to eliminate some threats to validity in individual studies.

And, multiple studies with different participants, investigators, and conditions increase the degree of confirmation that can be accorded to a particular theory.

TYPES OF VALIDITY
(Or, the multitude of sins that can be committed in doing research)

Campbell and Stanley (1963), Cozby (2001), and others classify validity as internal validity and external validity. Cook and Campbell (1979) added two additional types: statistical conclusion validity (often considered under internal validity) and construct validity of causes or effects (often considered under internal validity).

There is much debate about (a) which types of validity apply to which types of research, (b) the relative priority of types of validity, and (c) the interrelation of the types (Pedhazur & Schmelkin, 1991). To some degree, the types of validity that one considers applicable depend on the definitions used for each type of validity and the type of research attempted.

All of the group quantitative designs depend on the establishment of a relationship among the variables. In some designs (e.g., experimental) it is a causal relationship; in other designs (e.g., ex post facto) it is not.

And to some degree, all of the types of validity that we will discuss involve establishing the certainty of the relationship among the variables. Thus, in the broad sense, all of the types of validity may be considered applicable to all group quantitative designs. To a more limited degree, the four types of validity can be considered to be applicable to single subject designs.

Following are definitions of types of validity and lists of potential threats to each. The material is summarized from Chapter 2 of Cook and Campbell (1979) and Parker (1990).

Internal Validity:
Validity with which statements can be made about whether there is a relationship between the variables in the form in which the variables were manipulated or measured.

Statistical Conclusion Validity:
Certainty of inferences about presumed covariation of variables at specified alpha level and variances – or, in other words, relative probability that the results of the statistical tests are representative of actual relationships in the data.

Construct Validity:
Approximate validity with which we can make generalizations about higher-order constructs from research operations.

External Validity:
Certainty of generalizability across populations, persons, settings, times, etc.

Internal Validity: **Validity with which statements can be made about whether there is a relationship between the variables in the form in which the variables were manipulated or measured (Cook & Campbell, 1979).**

Internal validity may or may not relate to whether or not a causal relationship can be established.

Internal validity refers to our relative certainty that our outcomes resulted from what we did or what we tested (Tuckman, 1988).

Some Threats to Internal Validity

History	For example, an event other than the treatment that occurs between the pre- and post-test).
Maturation	Changes in the subject that occur naturally with time; the changes affect the subjects' performance on the dependent variable differentially.
Testing	Refers to the fact that pre-testing may sensitize subjects in ways that affect the posttest scores more for the pre-tested subjects than others.
Instrumentation	Refers to the deterioration or changes in the accuracy of devices or observers used to measure the dependent variable, e.g., observers forget their training.
Statistical Regression	Grouping on the basis of scores tends to be inaccurate; extreme scorers' measures tend to regress toward the group mean. One or two items answered differently more drastically affects extreme scorers than those scoring near the mean.
Mortality	The loss of subjects during research due to death, absence, etc. This is a particular problem if the treatment causes mortality; the treatment groups' posttest mean would be contaminated by mortality.
Selection	This occurs when subjects are assigned to treatment and control groups on a nonrandom basis. This results in the groups being different on many variables.
Interaction of Selection, Maturation, etc.	Any two of the above threats to internal validity may interact to restrict the validity of the design.

Some Ways to Minimize Threats to Internal Validity (or to Control Error Variance)

- Random assignment of subjects to treatment or control groups
- Holding extraneous variables constant or restricting their range
- Including extraneous variables in the design to measure their effects
- Employing methods of statistical control
- Matching subjects in the treatment and control groups on contaminating, extraneous variables (Parker, 1990)

Note that these methods of control are listed in their order of preference.

Statistical Conclusion Validity: **Certainty of inferences about presumed covariation of variables at specified alpha level and variances (Cook & Campbell, 1979) – or, in other words, relative probability that the results of the statistical tests are representative of actual relationships in the data.**

Some Threats to Statistical Conclusion Validity

Low statistical power (topic of earlier lecture)

Violated assumptions of statistical tests

Fishing and the error rate problem (five neon tetras in the fish tank)

Mistaken acceptance of null hypothesis

Reliability of measures

Reliability of treatment implementation (how much subjects learned, degree of program implementation)

Some Ways to Reduce Threats to Statistical Conclusion Validity

- Do a pre-analysis statistical power estimation and consider obtaining more participants, raising alpha, or using a more powerful statistical test to achieve higher power. Power of .80 is desirable (Cohen, 1988).
- Use alpha reduction procedures when running multiple comparisons.
- Be sure that the instruments that you use are reliable.
- Avoid using gain scores, or use them with appropriate caution. We will discuss the problems of gain scores in a future session. Ferguson and Takane (1989, pp. 474-475) and Cook and Campbell (1979, pp. 182-185) address the problems of gain scores.

- Be sure that treatments are fully implemented.
- Understand the assumptions that accompany the statistical tests that you are using and the consequences of their violation under various circumstances.
- Remember, we don't prove the null hypothesis; we only fail to reject it.

Construct Validity: "Approximate validity with which we can make generalizations about higher-order constructs from research operations" (Cook & Campbell, 1979, p. 38).

Some Threats to Construct Validity of Causes and Effects

Inadequate preoperational explication of constructs (e.g., the wrong bait)

Mono-operational bias (e.g., just one kind of fish net)

Evaluation apprehension (e.g., test anxiety)

Experimenter expectancies (e.g., the "Rosenthal effect")

Interaction of different treatments

Restricted generalization across constructs

Some Ways to Reduce Threats to Construct Validity

- Provide a good operational definition for the construct. Recall from our discussions of variables (i.e., a construct of interest) the importance of operational definitions and defined measurement scales. In other words, be specific in defining the construct of interest.
- Be sure to choose a dependent variable or outcome measure that really measures your intervention. For example, does the number of rehabilitation closures really measure the quality of service?
- Isolate the other constructs that can confound or confuse the issue
- USE MULTIPLE MEASURES AND MANIPULATIONS WHEN POSSIBLE (Cook & Campbell, 1979). Analogy: More than one type of fishnet, more than one type of fishing procedure.

External Validity: Certainty of generalizability across populations, persons, settings, times, etc. (Cook & Campbell, 1979; Tuckman, 1988).

Some Threats to External Validity

Interaction of Testing and Treatment – the pretest increases or decreases the respondents' responsiveness or sensitivity to the treatment; as a consequence the results are not generalizable to the nonpretested population from which the treatment group was selected.

Interaction of Selection and Treatment – research subjects are frequently volunteers or individuals who are prone to seek out research participation. Such persons may have traits that tend to enhance or diminish the effects of the treatment. Thus the results are not generalizable to the population of interest, which includes nonvolunteers.

Reactive Arrangements – the treatment employed in a study, particularly if administered in an artificial, laboratory setting, may not be identical to the treatment utilized in applied settings. Therefore, the results of the research may not be generalizable to the field (e.g., elephant).

The following threats to external validity are explained in Pedhazur and Schmelkin (1991).

Treatment-Attributes Interactions

Treatment-Setting Interactions

Posttest Sensitization

Some Ways to Reduce Threats to External Validity

- random selection
- deliberate sampling for heterogeneity
- using a "Hawthorne" control group
- examining potential interactions

Note

"The process of doing an experiment – that is, exercising some control over the environment--contributes to internal validity while producing some limitation in external validity." And ...

"External validity is of little value without some reasonable degree of internal validity" (Tuckman, 1988, p. 6).

The attached checklist can help when one is evaluating the threats to validity of a study.

VALIDITY CHECKLIST

Recall that these types are only illustrative. There are many more.

INTERNAL VALIDITY

__ history
__ maturation
__ testing
__ instrumentation
__ statistical regression
__ mortality
__ selection

STATISTICAL CONCLUSION VALIDITY

__ low statistical power
__ violated assumptions of statistical tests
__ fishing and error rate (e.g., multiple F-tests)
__ mistaken acceptance of null hypothesis
__ reliability of measures
__ reliability of treatment implementation

CONSTRUCT VALIDITY

__ inadequate preoperational explication of constructs
__ mono-operational bias
__ evaluation apprehension
__ experimenter expectancies
__ interaction of different treatments
__ restricted generalization across constructs

EXTERNAL VALIDITY

__ interaction of testing and treatment
__ interaction of selection of selection and treatment
__ reactive arrangements
__ treatment-attributes interactions
__ treatment-setting interactions
__ posttest sensitization

Evaluating Articles for Threats to Validity

The following lists are not exhaustive. Rather they illustrate some ways in which designs address threats to validity and some types of threats to validity that can remain.

Internal Validity

Internal validity is a focus if there is an intervention.

The research design can address threats to validity through

- random assignment of subjects to groups (experimental or control)
- holding extraneous variables constant or restricting their range (for example, focusing only on young adults)
- including extraneous variables in the design to measure their effects (e.g., including pre-test measures to see how pre-test levels influence effectiveness of the treatment)
- employing methods of statistical control (e.g., Analysis of Covariance [ANCOVA])
- matching subjects in the treatment and control groups on contaminating, extraneous variables

Threats to validity remain when any of the following happen any of the following situations, described in the lecture notes, occur:

history, maturation, testing, instrumentation, statistical regression, mortality, and selection

Statistical Conclusion Validity

Statistical conclusion validity is an issue whenever statistical tests are used to test hypotheses.

The research design can address threats to validity through

- considerations of statistical power
- alpha reduction procedures (e.g., Bonferoni technique) when multiple tests are used
- use of reliable instruments
- assurance of treatment implementation
- testing and abiding by the assumptions for valid use of statistical procedures
- never accept the null hypothesis (rather, we fail to reject it).

Threats to validity remain when

- there is low statistical power (this is only a threat if no significant differences are found)
- assumptions for statistical tests are violated
- multiple tests (more than 3) are conducted without alpha reduction
- a null hypothesis is falsely accepted

- measures with low reliability are used or the reliability for the sample is not reported
- reliability of treatment implementation is not assured

Construct Validity

Construct validity deals with what we are measuring and how we are measuring it.

The research design can address threats to construct validity through

- providing clear operational definitions of variables
- using multiple measures (e.g., multiple instruments) to evaluate the same construct
- using multiple ways of measuring a construct (e.g., behavioral observation and a standardized measure);
- using a measure that is free from confounding constructs and exactly measures the construct of interest (e.g., the right net)
- removing confounding constructs, like participant test anxiety and experimenter expectancies

Threats to validity remain when

- only one measure is used
- the measure is not clearly defined
- the construct of interest is only measured in one way
- confounding constructs are present (e.g., test anxiety, experimenter expectancies)

External Validity

External validity deals with the extent to which you can generalize results beyond the study sample.

The research design addresses external validity when

- the sample is randomly selected from a broad population
- the sample is deliberately selected for heterogeneity (e.g., multi-stage cluster sampling)
- a Hawthorne control group is used
- interactions of selection and treatment or testing and treatment are examined.

Threats to validity remain when

- volunteers or a purposive or convenience sample are used
- a pre-test interacts with the treatment
- the setting of the intervention is considerably different from the real world
- attribute treatment interactions are not addressed (e.g., people of different levels on the pre-test benefit differently from the intervention)

RELATIONSHIP OF VALIDITY TO RESEARCH DESIGN IN QUANTITATIVE RESEARCH

Research design is "the plan and structure of investigation so conceived as to obtain answers to research questions" (Kerlinger, 1986, p. 279). Its purposes are:

- to provide answers to research questions and
- to control variance (Kerlinger, 1986).

And, according to Cohen (1988), "Experimental [research] design is an area of inquiry wholly devoted to the removal of irrelevant sources of variability for the increase of precision and therefore the increase of the statistical power of tests of null hypotheses" (p. 8).

In other words, ***research design is a method of reducing the alternative explanations (i.e., rival hypotheses) related to a study.***

Recall that threats to validity are plausible rival hypotheses (i.e., other ways of explaining the results rather than the author's hypothesis).

Research design helps us to eliminate some threats to validity in individual studies.

And, multiple studies with different participants, investigators, and conditions increase the degree of confirmation that can be accorded to a particular theory.

Control

Control is a word that we encounter often in research articles and research texts. Like many words that we encounter in research, it has a variety of different meanings, which depend on the purposes of the researcher.

Control is the major tool used by research design to eliminate threats to validity (i.e., rival hypotheses).

Control has two purposes. They are:

1. ruling out valid threats to inference
2. adding precision, that is increasing the ability (i.e., statistical power) to detect small observed effects (Cook & Campbell, 1979)

Analogy: Decreasing the static on the phone line.

Pedhazur and Schmelkin (1991) identified four types of control: (a) manipulation, (b) elimination or inclusion, (c) statistical, and (d) randomization. Another way to categorize methods of control is:

- control of the situation, that is keeping out extraneous forces (e.g., testing students in the same room)

- control over the independent variable (i.e., the treatment) (for example, control of assignment of persons to different treatment groups, assurance of control over the nature and extent of the implementation of the treatment)
- controlling for an identified threat to validity through research design (for example, one can control for gender by only testing only one gender). Another example might involve measuring the effect of a particular program on changing participant's attitudes towards persons with disabilities. Pre-program attitudes could be controlled for by pre-testing and controlling for statistically controlling for pretest scores (e.g., through ANCOVA) when comparing the experimental and control groups on the post-test.

Control, in whatever way it is categorized or implemented, has the major purpose of removing threats to validity. Such threats can also be conceptualized as plausible rival hypotheses.

Putting it Together in Research

Parker (1990) correctly observed that "the perfectly designed study exists only in textbooks; in reality there is no such thing as flawless research" (p. 620).

Before you get discouraged with all the uncertainty, think of Monet. Each research study is like a brush stroke in the painting of a field of inquiry. One study doesn't determine the whole picture; it is but a brush stroke. Nonetheless, each brush stroke should be carefully planned.

Parker (1990) suggests a twofold approach for persons designing research.

> First, avoid errors by selecting appropriate research design and statistical analyses at the planning stage.
>
> Second, be aware of and report threats to validity.