

Evaluation of Experiments

Reliability of Measures

Consistency, repeatability

The major forms of reliability estimate:

1. Test - retest. Give the same test to the same people at some later time, and compute the correlation. Time lag varies.
2. Alternate forms. Two forms of the test (form A, form B) are developed and given at the same time to a group of people.
3. Internal consistency. One test is given once. The individual items themselves or else groups of items serve as their own alternate (mini) forms.

Split half- cut test in two, compute correlation. Then correct for short test with Spearman Brown Prophecy Formula.

Alpha - only one estimate. Theoretically estimates the expected correlation with a "randomly parallel" test.

Internal and External Validity

- both are about the quality of inference or meaning of the study.

Internal validity is about whether the study shows what it says - did the IV cause changes in the DV or is there some other explanation?

External validity is about whether the results apply to other situations that weren't part of the study, that is, whether the study results generalize to other situations or contexts.

There are many recognized threats to internal validity: He Made The IRS MIDday.

- History - meaningful events change people b/t repeated measures. Ex: job attitudes, strike, job attitudes; Group cohesion, win/lose ball game; Magic Johnson study.
- Maturation - Changes in people during a study. People get bored, fatigued, etc. Ex: physical ability measures; long survey.
- Testing. Practice effect. Reactivity to the test. Ex: attitude measures (e.g., beliefs about charitable contributions, film, attitude again); faculty opinions about anything.
- Instrumentation. Changes in the measuring instrument, e.g., wear on a work sample test. Computer breakdown. Human judges & double blind.
- Regression - people selected on the basis of test scores tend to have retest scores closer to the mean. Ex: people selected for remedial education; second year slump.
- Selection of group members - group composition is unequal before start of study. Ex: volunteers for effects of beer on driving; athletes vs. nonathletes (for differences in physical fitness on some dv or for differences in social networks on some dv).
- Mortality - people quit the study. Ex: influence of exercise on health and well being; techniques to quit smoking.
- Interaction with selection - although groups are equal at start, one group responds to something other than the IV (hardest to understand). Ex: Male and female groups

attitudes on aids before and after Magic Johnson announcement (selection & history).
College bound vs. voc tech on series of cognitive ability tests (selection & maturation).

- Diffusion (imitation of treatments) - People tell each other of the treatment. Ex: visualization in sports; website for class.

If given an example, be able to identify the term. If given a term, be able to identify an example.

Protecting Internal Validity

Randomization

Pre-post designs

Pretest-posttest control group design

Time ⇒

R	O1		O2
R	O3	X	O4

R = random assignment; O = measurement or observation; X = treatment.

Controls for history, maturation, testing and regression (so long as random assignment from extreme group). History, maturation, testing and regression effects can be seen because of pre-post comparisons. Can check on mortality. Instrumentation can be a problem. Allows us to see whether groups are equal at the start.

Solomon Four-Group Design.

Time ⇒

R	O1		O2
R	O3	X	O4
R			O5
R		X	O6

The is the same as the pretest-posttest control group design, but it adds two groups with no pretest. Because there is no pretest for the last two groups, we expect no Testing effects for those groups, and thus perhaps a better estimate of the actual effects of a single treatment. Problem with analyzing the data.

Posttest only control group design.

R			O5
R		X	O6

Because of the magic of random assignment, this is a good design.

External Validity

Here we want to generalize from our study to something beyond it. We almost never do a study to find out if X influences Y for a small number of people in single circumstance for a tiny period of time. Rather, we are interested in things that influence lots of people in many situations over a meaningful period of time.

Population Generalization- the study participants should represent some larger segment of people. Ex: The problem of studying only college sophomores. Ex: psychotherapy

studies with normals. This problem is solved by representative sampling from a population or by replicating the study with other populations.

Environmental Generalization - applying the results of the study to other situations.

Ex: studying managerial behaviors in the laboratory (e.g., performance appraisal, power & politics). Hard to know how to fix - depends on the threat (e.g., consequences to the actor, types of stimuli, etc.).

Temporal Generalization - do the study results apply over time? Ex: studies of the effects of clothing shown in pictures of people on hiring decisions. Solved by replication over time.