Internal Validity

One of the keys to understanding internal validity (IV) is the recognition that when it is associated with experimental research it refers both to how well the study was run (research design, operational definitions used, how variables were measured, what was/wasn't measured, etc.), and how confidently one can conclude that the change in the dependent variable was produced solely by the independent variable and not extraneous ones. In group experimental research, IV answers the question, "Was it really the treatment that caused the difference between the means/variances of the subjects in the control and experimental groups?" Similarly, in single-subject research (e.g., ABAB or multiple baseline), IV attempts to answer the question, "Do I really believe that it was my treatment that caused a change in the subject's behavior, or could it have been a result of some other factor?" In descriptive studies (correlational, etc.) internal validity refers only to the accuracy/quality of the study (e.g., how well the study was run-see beginning of this paragraph).

In their classic book on experimental research, Campbell and Stanley (1966) identify and discuss 8 types of extraneous variables that can, if not controlled, jeopardize an experiment's internal validity.

1. **History**--These are the unique experiences subjects have between the various measurements done in an experiment. These experiences function like extra, and unplanned, independent variables. Compounding this, the experiences are likely to vary across subjects which has a differential effect on the subjects' responses. Studies that take repeated measures on subjects over time are more likely to be affected by history variables than those that collect data in shorter time periods, or that do not use repeated measures.

2. **Maturation**--These are natural (rather than experimenter imposed) changes that occur as a result of the normal passage of time. For example, the more time that passes in a study the more likely subjects are to become tired and bored, more or less motivated as a function of hunger or thirst, older, etc. As Isaac and Michael (1971) point out, subjects may perform better or worse on a dependent variable not as a result of the independent variable but because they are older, more/less motivated, etc.

3. **Testing**--Many experiments pretest subjects to establish that all the subjects are starting the study at approximately the same level, etc. A consequence of pretesting programs/protocols is that they can contaminate/change the subjects' performance on later tests (e.g., those used as dependent variables) that measure the same domain beyond any effects caused by the treatment itself.

4. **Instrumentation**--Changing the measurement methods (or their administration) affects what is measured. Additionally, if human observers are used, it may be the judgment of the observer(s) that change over time rather than the subjects' performance.

5. **Statistical Regression**--When subjects in a study are selected as participants because they scored extremely high or extremely low on some measure of performance (e.g., a test, etc.), retesting of the subjects will almost always produce a different distribution of scores, and the average for this new distribution will be closer to the population's. For example, if the chosen subjects all had high scores initially, the group's average on the retest will be lower (i.e, less extreme) than it was originally. Conversely, if the group's mean was originally low, their retest mean would be higher.

6. **Selection**--The subjects in comparison (e.g., the control and experimental) groups should be functionally equivalent at the beginning of a study. If they are, then observed differences between the groups, as measured by the performance dependent variable(s), at the end of the study are more likely to be caused only by the independent variable instead of organismic ones. If the comparison groups are different from one another at the beginning of the study the results of the study are biased.
Internal and External Validity: General Issues

7. **Experimental Mortality**--Subjects drop out of studies. If one comparison group experiences a higher level of subject withdrawal/mortality than other groups, then observed differences between groups become questionable. Were the observed differences produced by the independent variable or by the different drop out rates? (Mortality is also a threat when drop out rates are similar across comparison groups but high.)

8. **Selection Interactions**--In some studies the selection method interacts with one or more of the other threats (described above), biasing the study's results.

External Validity

The extent to which a study's results (regardless of whether the study is descriptive or experimental) can be generalized/applied to other people or settings reflects its external validity. Typically, group research employing randomization will initially possess higher external validity than will studies (e.g., case studies and single-subject experimental research) that do not use random selection/assignment. Campbell and Stanley (cited in Isaac & Michael, 1971) have identified 4 factors that adversely affect a study's external validity.

1. An interaction between how the subjects were selected and the treatment (e.g., the independent variable) can occur. If subjects are not randomly selected from a population, then their particular demographic/organismic characteristics may bias their performance and the study's results may not be applicable to the population or to another group that more accurately represents the characteristics of the population.

2. Pretesting subjects in a study may cause them to react more/less strongly to the treatment than they would have had they not experienced the pretest. In such situations the researcher(s) cannot conclude that members of the population who were not pretested would perform in a similar manner to the participants in the study. Restated, to generalize the results of the study the researcher would have to specify that a particular type of pretesting also be done because the pretesting could be serving as an extra, unintentional independent variable.

3. The performance of subjects in some studies is more a product or reaction to the experimental setting (e.g., the situation where the study is conducted) than it is to the independent variable. For example, subjects who know they are participants in a study, or who are aware of being observed, etc., may react differently to the treatment than a subject who experienced the treatment but was not aware of being observed, etc.

4. Studies that use multiple treatments/interventions may have limited generalizability because the early treatments may have a cumulative effect on the subjects' performance. If a group experienced treatment X1, and the first treatment was followed by a second (X2), their measured performance after X2 will be affected by both treatments not just X2's because the effects of X1 are not erasable.

Increasing Internal and External Validity

In group research, the primary methods used to achieve internal and external validity are randomization, the use of a research design and statistical analysis that are appropriate to the types of data collected, and the question(s) the investigator(s) is trying to answer. Single-subject experimental studies almost always have high internal validity because subjects serve as their own controls but, as mentioned earlier, are extremely low with respect to external validity. Single-subject studies acquire external validity through the process of replication and extension (i.e., repeating the study in different settings, with a different subject, etc.). The results of group studies are also more acceptable by the scientific community when replicated.

References