A Beginner's Guide to Behavior Research Methods

By Lauren S. Seifert

Part I: Terms and Definitions

Data: A plural noun; Pieces of information that are collected in a systematic fashion. The singular form of this noun is "datum".

Empirical (an adjective): a term applied to research that involves measurement of observable behaviors or effects. This adjective refers to those measurement techniques that attempt to uncover universal truths or principles by way of objective, unbiased observations.

Ontology: a set of assumptions about whether absolute truth exists (objective ontology) or does not (subjective ontology).

Epistemology: a method for discovering things.

Grounded Theory: Perhaps, a term dating to the 1960's, grounded theories are derived *inductively* from data. Refer to Strauss and Corbin (1991) for a description of the systematic way that meaning can be derived from data in order to build theory.

Human Research: According to the Dept. of Health & Human Services, USA, it is collection of data about a person or persons in an effort to draw generalizable conclusions (paraphrased). See 45 CFR 46 of the US Federal Code of Regulations.

The Scientific Method: a general set of principles that is followed by traditional sciences (e.g., physics, biology, chemistry, sociology, psychology). The guiding ideals of the scientific method are that there are basic, underlying universal principles and that those principles/laws can be discovered via precise and unbiased measurement techniques. Such principles are objective truths. The Method listed below is a "long" version. For an abbreviated Method with comparisons to action research, see the Table that follows this list.

Dr. Seifert's Summary of the Scientific Method (in "steps"):

Step 1: Decide on a topic area.

Step 2: Define the problem through review of pertinent literature, and develop a research question.

Step 3: Define key variables and develop one or more predictions about them. These predictions are called "hypotheses."

Step 4: Design one or more research studies that will help you to answer your research question and address your predictions.

Step 5: Design specific research measures and pilot test them.

Step 6: Refine research measures based on the results

of your pilot testing.

Step 7: Run your study. Collect data.

Step 8: Code, compile, and analyze your research data into results that can be interpreted.

Step 9: Interpret research results and draw conclusions.

Step 10: Revisit your research hypotheses, build new predictions, and develop additional studies to address old/new research questions.

Comparing Traditional Science and Action Research

The Scientific Method	**an Action Research orientation**
1a) Identify a problem area.	1) Coghlan & Brannick's (2010) "Pre-Step": Defining Context and Purpose.
1b) Define the problem –a problem statement via objective analysis of the existing literature.	2) "Constructing" the problem space through dialogic activity among stakeholders.
2) State hypotheses/predictions.	2) continue constructing the problem space; and 3) Begin a plan for action.
3) Design an empirical study: one that uses objective (third-person) measures.	3) continue planning for action. Plans for action include personal (first-person), collaborative (second-person), and consultative (third-person) views.
4) Run the study/Collect Data.	4) Act. Note that actions are monitored and observed for their effects. The investigator: pays attention, acts intelligently, uses good judgment, and is practical.
5) Code* and analyze the data. * To code typically means to render data into a form that can be analyzed with statistics. 6) Draw conclusions from data analyses which relate back to theories and hypotheses.	5) Evaluate the action and its impact.

Defining Traditional Science and Action Research

Most of the research methods described in this document are of the conventional (or traditional) type. Their foundation is an objectivist ontology (i.e., assuming that there are absolute truths), and they rely on an objectivist epistemology (a way of knowing that seeks to measure truths without bringing in human/other biases).

Conventional science uses The Scientific Method. The hallmark of traditional research is the “true experiment”. A true experiment is a study with random selection from a population and random assignment of persons to research conditions or groups; in it at least one independent variable is manipulated and that variable must have at least 2 levels or conditions. True experiments are EXPLANATORY and allow one to draw conclusions about cause and effect.

Conventional science also uses DESCRIPTIVE methods, such as interviews, surveys, and quasi-experiments. It does so out of necessity, because some situations do not lend themselves to running true experiments. Descriptive research does not allow one to draw conclusions about cause and effect; it merely allows one to DESCRIBE relationships between variables (such as with correlations). A descriptive study cannot explain whether one variable is causing changes in another.

What is “action research”? It is an orientation that provides an alternative to conventional, positivist science. It has gained some popularity in such disciplines as education, the arts, business/management, social work, counseling, and nursing. An excellent guide to action research is authored by Coghlan and Brannick (2010). Like those in traditional research, action researchers assume an objectivist ontology (that there are absolute truths). However, an action researcher is generally less optimistic about the abilities of scientists to uncover truth. As a result, action researchers are much more likely than conventional scientists to use subjectivist epistemologies (ways of knowing that rely on people’s perceptions) and to focus on people’s experiences. Nonetheless, action research <u>is not phenomenology.</u> Action researchers assume that absolute truth exists…even if one cannot necessarily discern it. Phenomenologists, on the other hand, dwell in the subjective…having subjectivity as their ontology and their primary epistemology.

There are many approaches to action research, including psychologist Kurt Lewin’s classical approach whereby the action researcher usually acts as a consultant whose client needs assistance in making a decision, solving a problem, or managing an organizational change (see Coghlan & Brannick, 2010). In this approach, the consultant works together with the client through cycles of developing plans, acting based on those plans, and assessing those plans and actions in order to effect positive change. Classical action research is popular in business and in education.

Another type of action research is reflective practice: oft adopted by professionals who wish to engage in high degrees of scrutiny related to their own work (e.g., mental health providers, health providers). Schon’s work is frequently used as a guidepost for reflective practice. It focuses on one’s first-person experiences and aims to improve practice through one’s

improved understanding of his/her own professional planning and action (see Schon, 1987).

Another type among the many varieties of action research is participatory action research (PAR) (Coghlan & Brannick, 2010). It begins at the first-person level of inquiry, whereby the researcher is also an active participant in planning, change, and evaluation. However, unlike reflective practice, PAR usually involves a highly collaborative effort that is aimed at institutional or social change. There is an emphasis on cooperation and understanding among and between all stakeholders. Furthermore, there is a strong value of stakeholder equality so that all might be heard and play roles in effecting change.

The "Person" Level of Inquiry

It is vital to point out that different research orientations take various approaches to "the person". Traditional research removes itself from the person and tries to maintain objectivity by being at the "third-person" level of inquiry. This means that a conventional scientist must not have any personal ("second-person") ties to any of the research subjects ("first persons"). This "third-person" approach might help the researcher to be relatively free of biases, or so it is argued among traditional researchers.

Action research dwells more in the first- and second-person realms, although it does also take into account third-person data. As an example, let's consider a company that wishes to increase its client base. Top managers might first look at market trends (third-person data). Then, they might study their existing clients ("second persons"; i.e., persons with whom company employees interact) via surveys, interviews, or focus groups. In addition, company management might bring in a consultant (a "second person") to guide focus groups among company employees ("first persons"). By studying all of those constituencies, the consultant and the company managers might be able to chart a course for change, take actions to increase their client base, and then evaluate those actions for outcomes (whether positive or negative).

Important Terms Related to Sampling and Data Collection

Sampling (a verb): the acts of deciding who will be asked and then soliciting persons to take part in a study; if animals or objects are to be studied, then, the act of deciding which animals or objects will be scrutinized.

To "code" data (a verb): To convert pieces of information, which have been collected in a research study, to a form that can be more readily analyzed. Often, data coding involves transformation into quantities that can be analyzed with statistics.

Ethology: a somewhat antiquated term, it refers to the study of animal behavior.

Stimulus (plural, stimuli): simply, "a thing"; anything that has the potential to elicit a response.

Response: a behavior in reaction to one or more stimuli.

Variable: any of a number of factors that might be manipulated, controlled, and measured in a research study. Two basic types of variables are "Independent Variables" and "Dependent Variables."

Subject: the person, animal, or object being observed by a researcher.

Participant: considered a synonym of "subject" by some researchers, this term is generally considered to be more "politically sensitive" and is used to acknowledge a subject's right to choose to take part or to refuse to take part in a study. "Participant" is much more often used as a term in survey research—when those who take part are actively participating in giving data. The term "subject" is much more likely to be used in studies of objects, animals, children, or of adults whose behaviors are being actively manipulated (as in a true experiment).

Respondent: another common term for "participant" that is commonly used in survey research.

Population: a universal set; <u>all</u> persons, animals, or objects that share a predetermined set of characteristics. Examples: "All polar bears on the earth" is the population of polar bears. "All students currently enrolled at Malone University" is the population of Malone University students. A list of all members of a population is called a “sampling frame”.

Sample: a selective set; all persons, animals, objects or stimuli that has been selected for observation. Examples: 200 Malone University students who are approached as they walk out of the dining hall, who are asked to participate, and who agree to take part in a survey about their career goals; 30 polar bears who are observed in the Arctic during a scientist's exhibition there.

Research Design: usually a term applied to conventional research, it refers to the specific number and types of variables being manipulated and measured in a study. The description might include the number of levels/conditions of each independent variable. The research design type could be described numerically (e.g., a 2 X 2), with words (e.g., a within-subjects design), or with a combination (e.g., a 2 X 2 factorial design with an additional, nested within-subjects variable).

Sampling

Probabilistic Sampling (also "Probability Sampling"): a class of sampling types, all of which have the following characteristic: that the chance of any given member of the population being selected to take part IS KNOWN by the researcher. For example, in simple random sampling, each member of the population has an equal chance of being selected.

Non-Probabilistic Sampling (also "Non-Probability Sampling"): a class of sampling types all of which share the following trait: that the chance of any given member of the population being selected to take part IS NOT KNOWN by the researcher, e.g., convenience sampling.

Select Types of Sampling:

Simple Random Sampling: occurs when all members of a population are given an equal chance of being selected to take part in a research study. This is a "probabilistic" method of sampling.

Representative Sampling: attempts to utilize the techniques of random sampling, while adding an algorithm that attempts to bring in participants from population sub-groups in equal proportions to their representation in the population. Example: In sampling residents of New York City, a researcher uses census data to determine various sub-groups, like African-American persons, Asian-American persons, Latino-American persons, non-American citizens, etc. Then, the researcher attempts to build a study sample that includes those sub-groups in the same ratios that they represent in the city's actual population. This is a "probabilistic" method of sampling.

Convenience Sampling: occurs when a researcher is selective in sampling—selecting participants on the basis of their availability, proximity, and/or the ease by which they might be studied. This is a non-random method of sampling. This is a "non-probabilistic" technique.

Snowball Sampling: a type of convenience sampling that identifies members of a population that one wishes to study. Once those members are identified, they are asked to help identify other potential participants—typically, who are like them. Example: Finding participants for a survey about homelessness by visiting a shelter and asking homeless persons to tell their friends 'on the street' about the study. This is a "non-probabilistic" technique.

Sampling in "Waves": a basic term, this technique can be used with just about any sampling method. It consists of sampling at distinct points in time in order to increase the validity of one's sampling method. For example, a researcher who desires to collect data from a representative sample of Malone University undergraduates might begin by using a random number table applied to the student directory. S/he would call potential participants who had been randomly selected. This would be the "first wave." After conducting this first round of phone calls, the researcher would carefully assess which student sub-populations have been missed in the first wave (e.g., perhaps noticing that commuters had been reached in higher numbers than their actual proportion in the student population). A second wave of sampling might correct the imbalance.

Select Types of Research:

Naturalistic observation: a "non-experimental" research study; Generally speaking, there is no manipulation of variables by the researcher. S/he merely watches, listens, & observes behaviors and records information about them. In one type of naturalistic observation, called "observation of public behavior" the researcher watches and/or listens to behaviors that occur in public. In “participant observation” the researcher takes part and documents his/her experiences and observations as an “insider”. A DESCRIPTIVE type of research.

Case study: the behaviors of one person or of a small group are studied. Case studies can be non-experimental, quasi-experimental, or experimental—depending upon how the researcher manipulates (or does not manipulate) the research environment. DESCRIPTIVE RESEARCH

Survey: any of a number of DESCRIPTIVE techniques, including paper-&-pencil questionnaires, face-to-face interviews, phone interviews, and focus group interviews, online surveys, and computerized balloting. Surveys aim to collect information directly from persons by asking for their responses to questions or reactions to items. See "Selected advice for composing better surveys" and "Basic types of survey items" below.

Quasi-experiment (also, "pseudo-experiment"): like a true experiment, this type of research study utilizes naturally occurring variables, like age, weight, height, gender, etc., to categorize subjects. Then, the potential effects of naturally occurring variables on other behaviors/traits (like depression, spending behavior, reading aptitude, spirituality, general health) might be studied. Quasi-experiments include less overall control than do true experiments. DESCRIPTIVE

True experiment: a researcher controls and manipulates various aspects of the environment (called Independent Variables) in order to find out how it will affect a person's or an animal's behavior (the behavior that is influenced is called the "Dependent Variable"). EXPLANATORY

Research Involving 2 or More Techniques

Mixed Methods: a label reserved for studies that utilize BOTH qualitative and quantitative methods during data collection (e.g., measuring someone’s blood pressure [numeric] and asking the person whether s/he feels stressed [qualitative]) (See Creswell, 2014).

Multi-Method: different than “mixed methods”. This term is BROAD, referring to any study with 2 or more approaches to data collection and/or data analysis.

Terms Related to the “Repeatability” and “Meaningfulness” of Data

Reliability: a general term, this refers to the likelihood that a specific result can be repeated or observed again.

Replication: a repeated study (often conducted in order to determine the reliability of research results).

Limited replication: a repeated study in which some conditions or variables are altered from the original study. Often, such replications attempt to repeat a study "with improvements."

Conceptual replication: the theory or model that underpins a study's design is used to devise another method for testing the phenomenon in question. In this type of replication, the methods and/or research design might be quite different than were those in the original study. Yet, the goal is to find further evidence for the phenomenon and—often—to find further support for the foundational theory or model.

Converging evidence, or converging operations: (after Garner, Hake, & Eriksen, 1956) this term refers to the attempts of researchers to determine underlying principles or truths about something by using several different research methods or measures in order to *converge* on those truths.

Validity: often confused with "reliability" this is a different term that refers to the general ability of a research study to uncover truth. When applied to a specific measurement technique, its validity is its ability to measure what it is designed to measure. Thus, validity pertains to the meaningfulness of a study’s manipulation and its results.

Face Validity: whether a measurement technique appears to make sense, whether it appears to measure that which it is designed to measure. Counterexample: As a measure of intelligence, one's hair color appears to have *poor* face validity.

***

Construct Validity: the degree to which something can be concluded about a higher-level or more abstract concept from an objective measure. For example, how much about a student's **knowledge of calculus** (a higher-level construct) can be concluded from his/her final exam score in a calculus class (with the exam performance being a direct, objective measure)?

Content and Criterion-based Validity are often assessed during an examination of a test's construct validity. They are:

Content Validity: whether the measure adequately covers the breadth of content that is to be tested. For example, if an employer wants to be sure that his/her accountants have adequate knowledge to accomplish all the tasks of their job, then the employer will want to be sure to assess all aspects of their accounting skills sufficiently to judge whether s/he can do the job.

Criterion-based Validity: Also, called Criterion-related Validity--a standard ("criterion") is established for measuring a phenomenon. Then, new measures are compared to the standard in order to verify their value as measures of the construct. For example, a new personality test might be compared to the NEO-PI in order to verify (or "validate") the new test's ability to measure the Big Five Factors of personality.

***

External Validity: the extent to which a conceptual model predicts "real-world" phenomena (see Cook & Campbell, 1979). Also used to refer to the extent to which the results of a research study reflect real-world phenomena.

Ecological Validity: thematically related to "external validity", it refers to the ways in which a research study and its results resemble real-world principles, environments, and occurrences (see Bronfenbrenner, 1977).

Internal Validity (related to "control"): the degree to which the results of a study are actually due to the researcher's manipulation of variable(s); the extent to which changes in the dependent variable(s) are owing to the manipulation of one or more independent variables.

Conceptualizing "Control" and "Generalizability" in Research:

A Graphic Representation of a "Continuum of Control"

LOW EXPERIMENTAL CONTROL	HIGH EXPERIMENTAL CONTROL
HIGHLY GENERALIZABLE	LESS GENERALIZABLE
HIGH EXTERNAL VALIDITY	LOWER EXTERNAL VALIDITY
LOW INTERNAL VALIDITY	HIGHER INTERNAL VALIDITY

I--I---I

Naturalistic Observation   Surveys & Quasi-Experiments   True Experiments

Part II: Illustrating Research Types, Designs, and Techniques

Observational Research (e.g., Naturalistic Observation)

Selected methods for sampling behaviors:

***By subject or case:** An individual is identified and his/her behavior is documented.

***By interval:** A time span is specified (e.g., 8:00AM – 10:00AM, daily) and only the behaviors occurring during that interval are documented.

***By duration:** A specified behavior (e.g., grooming in albino Sprague-Dawley rats) is observed and its duration is documented (e.g., 10 sec grooming; 30 sec grooming; 40 sec grooming; etc.).

***By frequency:** Generally, documenting with "hash" marks, the number of times a specific behavior occurs.

***By behavior category:** The researcher documents all categories of behavior that are observed, e.g., grooming occurred, drinking occurred, mating occurred. Verbal/textual descriptions of the behavior categories are usually given in this type of observation.

Surveys

Selected types of survey research:

***By paper-and-pencil questionnaire:** Respondents are asked to complete a paper version of the survey and return it to the researcher.

***By computer form or web-survey:** Respondents are asked to complete a computer form that is on a dedicated computer (e.g., the researcher's laboratory computer) or that is on-line (e.g., a web survey through "Survey Monkey" that is accessible via the internet).

***By phone:** Phone surveys usually involve a phone call from one person to another (researcher to respondent or vice versa) for the purpose of conducting a survey by voice, by phone.

***By pager, email, and/or "text messaging":** More recently, some researchers conduct simple surveys by page, text, email, or some other electronic device (e.g., of patient compliance by paging the respondent and asking, "It's 10AM. Have U taken UR medication?). These allow researchers to obtain snippets of critical information about behaviors, *in vivo*, i.e., as they occur in everyday life. A recent variant of this type of “survey” sampling can be done non-invasively with an electronic wrist band (like a “Fit-Bit) which can sample bio-statistics (e.g., heart rate, blood pressure) or even ask a participant to respond to a fixed-choice question (e.g., “Have you smoked a cigarette within the past hour?).

***By face-to-face interview:** An interviewer asks for responses to items/questions that s/he reads aloud or gestures (e.g., in ASL) to the respondent.

Basic types of survey items:

1) Open-ended: The participant can fashion the format of his/her response by writing, drawing, etc. Open-ended items often lead to richer data, but they are generally more difficult to "code." Coding open-ended responses so that they can be analyzed is often time consuming and difficult.

2) Closed-ended (or "fixed" or "multiple" choice): The survey item specifies response choices and, usually, asks the respondent to select one, rank the options, or select a fixed sub-set of options as his/her preferred choices.

Examples of Open- and Closed-ended Survey Items about University Student Sleep:

1) Open-ended:

Please, describe your typical sleep pattern. How well do you sleep and how often? (Please, write in the space provided.)

2) Closed-ended:

*Using the following scale, how would you describe your nightly sleep?

1	2	3	4	5
very irregular	irregular	somewhat regular	regular	very regular

*Of the following, which one BEST describes the <u>quality</u> of your nightly sleep pattern? (Please, circle one response.)

A. very satisfying
B. satisfying
C. neutral/unsure
D. dissatisfying
E. very dissatisfying

*Of the following, which one BEST describes the <u>amount</u> of your nightly sleep? (Please, circle one response.)

A. Far too little
B. Too little
C. Enough
D. Too much
E. Far too much

Selected advice for composing better surveys:

1) Use simple language. To assume no better than a 6th-grade reading level will provide comprehension of survey items by a broad, adult audience.

2) Ask about one issue per item. Don't combine issues within a single question or item. This can confuse respondents and muddle the data.

3) Use understandable response options and be sure that options are mutually exclusive, yet exhaustive.

4) Pilot test your draft survey with persons who are like the persons you will be surveying. This consists of giving the draft survey to a few people who can help identify problems with wording, clarity, length, duration, and subject matter.

5) Consider the ways you will code data. For example, in the second and third samples of closed-ended survey items (above), it might have been more helpful to code the response options numerically (e.g., "1" = far to little sleep, to "5" = far to much sleep). That numeric response scale would lend itself more easily to a statistical analysis of the respondents' perceived amount of nightly sleep.

6) Be wary of sexist, racist, or otherwise prejudicial language. If you would like to describe a behavior vignette or scenario, be sure your language is inclusive, so that it does not alienate or ostracize research participants. [For example: Notice the use of "his/her" in the instructions to subjects, below. This is appropriately non-sexist language in the instructions of an open-ended, vignette-type survey item.]

EXAMPLE: Please, read the following story and decide whether the person involved was justified in his/her behavior.

"Nino has studied all night for the big statistics exam. Upon reaching the classroom, he noticed that it was empty. Marta, another student in the class, stood nearby the door and asked, "Where is everyone?" It was 5-minutes before the class hour and Nino and Marta were the only persons at the classroom. As time ticked by, nobody else arrived. Finally, Marta said, "I'm going to call the professor's office, because something is just not right!" She called the office and while she spoke to someone on the other end of the phone, Nino said, "Oh, I'm just gonna go. We're obviously not having class or the test today." And Nino walked away.

Given that Nino and Marta each expected to come to the classroom and take an exam on that day and at that time, how would you explain their different behaviors?

Quasi-Experiments

Quasi- or pseudo-experiments are like true experiments, but they are not authentically “experimental” in nature. In a quasi-experiment, one or more independent variables (IVs) cannot be manipulated by the researcher; instead, at least one IV is chosen from among pre-existing characteristics (such as the subject traits of age, gender, race, ethnicity). Thus, the terminology is "pseudo" or "quasi"—to resemble a thing, yet *not* to be that thing.
Quasi- or pseudo-experiments seem like true experiments, but lack true control of independent variables. Thus, one might utilize designs that are *like* the designs of true experiments.

Independent variable (Abbreviated IV): things the researcher manipulates in a study in order to produce change. In quasi-experiments, some IVs cannot be manipulated. They are "pseudo" IV's (like a subject’s age or gender) but are often treated as if they were true IVs.

Dependent variable (Abbreviated DV): those things that are affected by independent variables. A researcher measures them in order to find out whether they are influenced by the IV.

True Experiments

As mentioned in a previous section, a true experiment is a study with random selection of subjects from a population and random assignment of persons to research conditions or groups. At least one independent variable is manipulated and that variable must have at least 2 levels or conditions. The levels or conditions of the independent variable might be manipulated “within subjects” (with each subject receiving all conditions/levels, but in a different order) or manipulated “between subjects” (with each subject being assigned at random to receive one condition/level of the independent variable). True experiments are EXPLANATORY and allow one to draw conclusions about cause and effect.

Independent variable (Abbreviated IV): things the researcher manipulates in a study in order to produce change.

Dependent variable (Abbreviated DV): those things that are affected by independent variables. A researcher measures them in order to find out whether they are influenced by the IV.

Hypothesis: an informed prediction about what will happen in a research study. Example: "Caffeine will increase anxiety among students who take caffeine."

Null hypothesis: a prediction stated as "null." In science, a researcher assumes that s/he can NEVER prove something with 100% certainty. Instead, a scientist must be humble. Always assume that there is error in research methods. Because of this error, state predictions as if they will not be proven.

Example: If I believe that caffeine will induce anxiety (as above), then I state the hypothesis as null, as follows, that "Caffeine will not change the anxiety levels of students who ingest it."

Statistical rejection of the null hypothesis: In the analysis of data from my research study, I will desire to **reject the null hypothesis** and show that my experiment supports the idea that caffeine does affect students' anxiety levels.

Examples of Quasi-Experiments and True Experiments

Quasi-experiment: An example with between-subject grouping according to a subject variable

In a study about memory and Alzheimer's disease, a researcher might utilize a between-subjects design in order to assess the influence of extra testing on participant performance. Performance at the second (or "real") test is compared across the two groups in order to assess the effects of a practice test.

GROUP (“pseudo” IV)	**held constant**	**Post-Test DV**
with probable Alzheimer's disease	study, then “practice” test	second, "real" test
without Alzheimer's disease	study, then “practice” test	second, "real" test

Notice that the experimenter can control study and test, and that s/he holds that constant at one level. That is, both groups get study time, followed by a practice test. Here, the IV of interest is "group" (with Alzheimer's versus without Alzheimer's). However, the experimenter has no control over who is assigned to groups. Persons come into the study with the disease or without it. Presence or absence of the disease is a “subject variable”. Thus, the researcher has no true control over the IV of interest. It is a "quasi" manipulation, using subject characteristics to create the between-groups comparison.

Another example of a Quasi-Experiment: Studying peer support completely within-subjects

Subjects	**Time 1 (DV)**	**Time 2**	**Time 3 (DV)**
Persons who choose to join a peer support group	**Survey about person’s well-being the first time they attend a meeting**	**Meeting attendance across several months**	**Survey about person’s well-being after six months of attendance**

The above design is sometimes called an intervention or program assessment. It is quasi-experimental because the peer support group attendees have self-selected into the program. The researcher <u>did not</u> randomly select them from the population and then assign them at random to peer support or no peer support. The IV (peer support) has only one level/condition. Thus, there is no true manipulation of an IV.

Experimental Designs

An Example: Manipulating caffeine entirely between subjects

Randomly select 40 students from among those who attend Malone University. Then, randomly assign the students to the "No Caffeine" and "Caffeine" groups. Use a survey about anxiety to evaluate students' anxiety levels.

The research design:

A simple between-groups design with one independent variable (also called a post-test only control and experimental groups design)

(NOTE: the IV has 2 "levels" of caffeine intake: no caffeine versus some caffeine. Each group receives just one level of the IV. Thus, the IV is being manipulated "between subjects".)

Group	IV	"Post-Test" DV
Control Group	water w/o caffeine	anxiety questionnaire
Experimental Group	water with caffeine	(same) anxiety questionnaire

Note that many researchers consider the foregoing design to be "quasi-experimental" because no attempt is made to establish the equivalence of the groups before the IV is manipulated. Any difference between groups on the post-test (anxiety questionnaire) might be due to the manipulation of the IV. Unfortunately, it might also be due to a lack of equivalence between groups (differences in anxiety) prior to testing. In order to try to reduce the chance of such a problem, participants should be assigned at random to groups. In order to be improved, this design can be turned into one that is readily accepted as a "true experiment"...

The pre-test and post-test control and experimental groups design (after Sekaran & Bougie, 2013)

Group	"Pre-Test" DV	IV	"Post-Test" DV
Control	anxiety questionnaire	water w/o caffeine	anxiety questionnaire
Experimental	(same) anxiety questionnaire	water with caffeine	(same) anxiety questionnaire

Participants are sampled at random from the population and then assigned at random to conditions. Each participant's anxiety level is measured twice: once at the start of the study (before any IV) and once after the IV manipulation. Here then, a researcher should be able to discriminate between pre-existing differences between groups and

a difference that is truly related to the manipulation of the IV. Thus, it's a true experiment.

Another research design:

A simple within-subjects design with one independent variable (also called a pre-test, post-test experimental group only design; also called a simple 1 IV repeated measured design).

This design is similar to the one with assessment of peer support (above). However, it is different because the researcher has control over the manipulation of the caffeine (but did not have control over the peer support intervention and who took part).

[NOTE: the IV has 2 "levels" of caffeine intake: no caffeine versus some caffeine. Unlike the between-subjects design (immediately above) all subjects in this within-subjects study (table below) receive all levels of the IV]

TIME	IV	DV
1 (all subjects at Time 1)	"Baseline" = no caffeine	anxiety questionnaire
2 (all subjects at Time 2)	"Experimental" = caffeine	(same) anxiety questionnaire

Note that the foregoing design is considered to be quasi-experimental by many researchers, because there is a "confound" between the order of levels of the IV and time. For whatever reason, the researcher has not controlled or manipulated time/order. To reduce the chance that an effect is due merely to the passing of time, this study can be improved (as below).

A repeated measures, counter-balanced design:

Here, one might use any number of methods to be sure that the order of baseline and experimental conditions is not the cause of any changes in the DV. One technique is to assign each subject to a random ordering of the two levels of the DV. Unfortunately, this can lead to the unintended consequence of one order (e.g., baseline-experimental condition) being present more often than another order (experimental condition-baseline). A better way to go is to "counter-balance" the design...making sure that all orders of the levels of the IV are present and that equal numbers of participants are assigned (at random) to each.

In a simple repeated measures experiment with just two levels of the IV, it is easy enough to counter-balance, because there are only two possible orders (i.e., baseline-experimental versus experimental-baseline). Here is what the counter-balanced design looks like represented in a tabular format.

Group	**Time 1**	**Time 2**
Order B-E	**"Baseline" (no caffeine) with anxiety questionnaire**	**"Experimental condition" (with caffeine) with anxiety questionnaire**
Order E-B	**"Experimental condition" (with caffeine) with anxiety questionnaire**	**"Baseline" (no caffeine) with anxiety questionnaire**

Once one does this, one has actually created a grouping variable, such that the "completely within-subjects design" now has a "between-subjects" IV added to it: order of conditions. Thus, this can now be called a "mixed design" with one repeated "within-subjects" IV (caffeine v. no caffeine) and one "between-subjects" IV (order of conditions).

Amazingly, even though the above design takes care of some possible confounds, there is STILL the chance that differences between groups BEFORE the study might end up being the reason that a group difference is found. There are a number of ways to handle the issue. One is to screen all participants before Time 1, but this would mean that all subjects would then receive the anxiety survey THREE TIMES. Repeated testing might be problematic, as one's experience with the survey can change the way s/he answers it on a subsequent occasion.

Another remedy is to manipulate all IVs between subjects, removing the within-subjects variable.

Another option is to retain the within-subjects manipulation as a variable that is "nested" within the between subjects design.

Both are illustrated in the following tables.

A "Solomon 4-Group" Design with 2 Between-Subjects Variables (removing the within-subjects order variable from above)

Group	Time 1	Time 2	Time 3
Control A	Nothing	Baseline (e.g., water with no caffeine)	Post-Test (DV) anxiety questionnaire
Experimental A	Nothing	Experimental (water with caffeine)	Post-Test (DV) anxiety questionnaire
Control B	Pre-Test (DV) anxiety questionnaire	Baseline (e.g., water with no caffeine)	Post-Test (DV) anxiety questionnaire
Experimental B	Pre-Test (DV) anxiety questionnaire	Experimental (water with caffeine)	Post-Test (DV) anxiety questionnaire

This example of a Solomon 4-Group Design can also be called a 2 X 2 factorial design, because it has two IVs that are manipulated BETWEEN subjects. Consequently, the IVs can be labeled as "factors". The two IVs are represented by the terms in the "multiplication problem".

The first term "2" signifies the between-subjects variable to manipulate pre-test (i.e., Pre-Test v. No Pre-Test; see Time 1). It has 2 conditions/levels. Therefore, the numeral "2" is used in the equation.

The second term in the equation denotes the between-subjects variable: caffeine or no caffeine. The "2" tells one that this IV has two levels/conditions, as well.

Just about any research design can be described as an equation, if it has 2 or more IVs (e.g., 3 X 2 X 2: a study with 3 IVs, one of which has 3 levels/conditions, while the other two IVs each has 2 levels/conditions).

Notes about factorial designs:

(1) Any design with a variable that is not received by all participants in all groups is called "nested". That is, an IV that is not combined (or "crossed") with all levels/conditions of all other IVs must be said to be embedded or "nested" within some other IV.

(2) Researchers rarely use 3-term designs; 2-term designs are much more common, because once there is more than one IV, a researcher must look for interactions between variables, and those can be very challenging to interpret. See Hall (1998) for more about factorial designs.

A "Solomon 4-Group" Design with 2 Between-Subjects Variables and with one Within-Subjects Variable that is nested within Level/Condition B of the Pre-Test v. No Pre-Test IV (see Time 3).

Group	Time 1	Time 2	Time 3
Control A	Nothing	Baseline (water with no caffeine)	Post-Test (DV) anxiety questionnaire
Experimental A	Nothing	Experimental (water with caffeine)	Post-Test (DV) anxiety questionnaire
Control B	Pre-Test (DV) anxiety questionnaire	Baseline (water with no caffeine)	Experimental (water with caffeine) Post-Test (DV) anxiety questionnaire
Experimental B	Pre-Test (DV) anxiety questionnaire	Experimental (water with caffeine)	Baseline (water w/o caffeine) Post-Test (DV) anxiety questionnaire

Unless one had a specific reason to believe that the within-subjects manipulation were needed, the first of the two Solomon 4-Group Designs would be preferable. The problem with the second Solomon 4-Group Design is that it confounds The Time 1 manipulation (a Between-Subjects manipulation of Pre-Test v. No Pre-Test) with the Caffeine v. No Caffeine intervention (now manipulated between-subjects for A Groups and within-subjects for B Groups).

Error in Research

No experiment is infallible. The above examples show a steady progression of design improvements, which should help with threats to the internal validity of a study.

Unfortunately, all studies are susceptible to potential biases from "mortality" (loss of subjects from a study). When subjects opt out of survey items or exit an experiment prior to its end, they might leave the relative representations of the population unequal across conditions or groups. Even groups that were equivalent on anxiety at the start of the above study might be different at Time 3, if participants are lost from one group more than they are lost from other groups.

Artifact: a global term for an error in research measurement, method, computation, or technique. Artifacts can originate from numerous sources, such as the ones listed below.

Some sources of ERROR in research:

Attrition: loss of subjects from a study during the interval over which they are studied. In longitudinal studies, participants might be lost in-between test sessions. In cross-sectional or single-time experiments, participants might be lost during the test session. Sources of subject loss can include loss due to: Subject election to leave/not to continue; subject illness or death; subject inability to return to the test site; etc. Loss of subjects is generally non-random. Thus, it can contribute to errors in the results & conclusions from a study. Loss of subjects due to death is a type of attrition called “mortality”.

Experimenter bias: the researcher's own imperfection leads him/her to perceive, record, and/or analyze data in a manner that introduces error. Example: reading a scale incorrectly.

Participant bias: a subject's own imperfection leads him/her to answer a question or perceive a situation in ways that lead to mistaken behavior. Example: a participant in a phone survey doesn't understand a question in a survey and answers in a way that does not represent his/her true opinion.

Effects of history: life events occur that change a study's results in ways that do not represent truth. Example: US residents born in 1925 appear to have greater intellectual strengths in math than do those born in 1945. The result is more likely due to the greater emphasis on math skills in grade-school from 1930-1935 than from 1950-1955. The result is probably NOT due to a difference in innate intellectual abilities. It is a result "created" by history.

Order or “Carryover” Influences: When two or more events happen within a study, there is always a chance that the earlier events can influence subjects’ responses to later events. Thus, the order in which the events occur (e.g., whether there is a pre-test, whether one survey happens before another survey) can impact a study’s results (see Elmes, Kantowitz, & Roediger, 2006).

Sampling bias: something has gone wrong in a researcher's sampling method and it has created the research result.

Demand characteristics: features of a study that cue participants about how to act or about how the researcher "wants" them to act.

Confounds: uncontrolled variables that affect the dependent variable, thereby making it seem as if the independent variable has produced the effect. Indeed, careful study of and control for potential confounds is critical for sound assessment of an IV's effect on the DV.

Hawthorne Effect: historically, thought of as the tendencies of research subjects to behave differently, merely because they are being observed.

Mixed Design: terminology that usually refers to the mixing of within- and between- subjects IVs within a single study.

In experimental research, a null hypothesis is set up as a prediction to be refuted by statistical analysis of the data. Because human researchers, research participants, and the experiment process are all imperfect, error can enter research and lead to erroneous conclusions. The following diagram reflects ways error can lead to false conclusions about the real-world (as in Type I and Type II errors).

Conditions in the Real World ("The Truth")

Experimental Results	**NULL IS TRUE**	**NULL IS FALSE**
Null Result (Researcher's Conclusion: The null hypothesis is accepted, and the researcher concludes that the IV does not affect the DV.)	**Correct acceptance of the null hypothesis. The IV manipulation does not cause a change in the DV.**	**INCORRECT acceptance of the null hypothesis. This is called a TYPE II ERROR.**
Null rejection (Researcher's Conclusion: The null hypothesis is rejected, and the researcher concludes that the IV does affect the DV.)	**INCORRECT rejection of the null hypothesis. This is called a TYPE I ERROR.**	**Correct rejection of the null hypothesis. The IV manipulation does cause a change in the DV.**

Recommended Reading

American Psychological Association. (1982). *Ethical principles in the conduct of research with human participants.* Washington, DC: Author.

American Psychological Association. (2009). *Publication manual of the American Psychological*

Association (6th ed.). Washington, DC: Author.

Baron, R.M., & Kenny, D.A. (1986). The moderator-to-mediator variable relationship in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Bordens, K.S., & Abbott, B.B. (2005). *Research design and methods: A process Approach* (6th ed.). New York, NY: McGraw-Hill.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32,* 513-531.

Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research.* Boston, MA: Houghton Mifflin.

Christensen, L.B. (2007). *Experimental methodology* (10th ed.). Boston, MA: Pearson.

Coghlan, D., & Brannick, T. (2010). *Doing action research in your own organization* (3rd ed.). Los Angeles, CA: Sage.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design & analysis issues for field settings.* Boston, MA: Houghton Mifflin.

Creswell, J.W. (2014). *A concise introduction to mixed methods research.* Thousand Oaks, CA: Sage.

Elmes, D.G., Kantowitz, B.H., & Roediger, H.L. (2006). *Research methods in psychology* (8th ed.). Belmont, CA: Thomson Higher Education.

Garner, W.R., Hake, H.W., & Eriksen, C.W. (1956). Operationism and the concept of perception. *Psychological Review, 63,* 149-159.

Grimm, L.G., & Yarnold, P.R. (Eds.). (1995). *Reading and understanding multivariate statistics.* Washington, DC: American Psychological Association.

Hall, R. (1998). *Psychology world: 2 X 2 between subjects factorial design.* Retrieved from https://web.mst.edu/~psyworld/between_subjects.htm

Kuhn, T.S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.

Lynd-Stevenson, R.M. (2007). Concerns regarding the traditional paradigm for causal research: The unified paradigm and causal research in scientific psychology. *Review of General Psychology, 11,* 286-304.

Popper, K.R. (1968). *The logic of scientific discovery*. London, UK: Hutchinson.

Schon, D. (1987). *Educating the reflective practitioner*. San Francisco, CA: Jossey-Bass.

Sekaran, U., & Bougie, R. (2013). *Research methods for business: A skill building approach* (6th ed.). West Sussex, UK: Wiley.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory and procedures.* London, UK: Sage.