11. Research Design

Over the course of history humans have evolved many different ways of understanding the world. **Epistemology** is the branch of philosophy that deals with the nature and foundation of knowledge. In Plato's dialogue *Theaetetus* knowledge is said to be **justified true belief.** Thus there are three attributes of knowledge: one believes a statement is true, it really is true, and the belief that it is true is justified.

The question of **justification** then becomes central to any theory of knowledge. Research is a particular theory of knowledge based on **empiricism**, i.e., that theories are tested against reality and are accepted or rejected depending on how well they correspond to evidence.

Empiricism is simply one of many different ways of justifying knowledge. Some others include:

- *An appeal to authority.*
- *Direct observation of the senses (naive empiricism).*
- *Mysticism.*
- *Logic or rationality.*

Scientists will sometimes use non-empirical methods. Scientists rely on the prior work of other scientists, which is one kind of appeal to authority. The very belief that there is an external world in which cause-and-effect are real phenomena might be thought of as a mystical belief. In fact, in some settings, such as the sub-atomic world described by quantum mechanics, this phenomenological approach fails to pass the test of correspondence to the facts.

However, the basic premise of science, that theories are accepted or rejected depending on how well they correspond to the facts, remains at the heart of scientific inquiry.

Most research projects pass through four basic phases:

A. Organization and Planning Phase
B. Data Collection Phase
C. Data Analysis Phase
D. Reporting Phase

The balance of these notes describes these phases in greater detail.

11.1. Steps in Phase A, Organization and Planning.

A.1. Background Research
A.2. Define the terms
A.3. Define the variables
A.4. Develop a model
A.5. Determine Objectives
A.6. Design measures
A.7. Select the strategy
A.8. Design the instruments
A.9. Design the Sample
A.10. Develop the budget

The steps are not necessarily done in the above order and often decisions in one step force the researcher to revisit decisions made in earlier steps. Further there are many other ways of dividing up the preliminary design steps: no one methodology can apply to all situations. The above taxonomy is provided as one of many possible templates.

A.1. Background Research.

Background research consists of learning what is already known about your research topic. Most generally this will consist of building an annotated bibliography of research articles already published on your topic. There are many tools available today to assist in bibliographic searches. A reference librarian can assist you in learning these tools (but you should not ask a librarian to do your search for you). Once you have constructed a list of articles, the next step is to read as many of them as you can, taking notes on the content. Those that are relevant to your research topic, along with your summaries, become your annotated bibliography. The bibliographic entries should conform to an accepted standard, such as the APA standard. In this way others can also find your bibliography useful, plus using the standard will assure that you have enough information to relocate the document.

A.2. Definitions.

The basic terms needed to carry out the research must be clearly defined. These definitions must be

- appropriate to what is being studied; and
- applicable in field research.

In particular, the person collecting data must be able to use the definition to decide whether or not it applies to whatever is being observed.

Another critical definition is the population that is being studied. This must also be clear, appropriate and applicable in field research.

In defining the population there are often two conflicting goals.

- Make the study as broadly applicable as possible.
- Make the study narrowly focused on a particular group or question.

The former goal leads to a different, and more expensive, kind of population than does the latter goal. Sometimes, even with sufficient resources, it is more effective to conduct several narrowly focused studies rather than one study on a large and complex population.

We will discuss this in greater detail in the section on sample design.

A.3. Variables.

A **variable** is any attribute that varies in quantity or quality from subject to subject.

Examples:

- Age
- Gender
- Sexual Orientation
- Weight
- Shoe Size
- IQ
- Hair Color
- Income

Selection of variables will depend on their relevance to the topic

Which variables you choose to measure is often driven by a theory about which variables matter. For example, in a study of social behaviors one would not ordinarily include eye color or hair color as variables. However, if the study dealt with the role of serotonin in social behaviors, one might include both eye color and hair color since natural serotonin levels are different in certain blue-eyed, blonde-haired populations.
Variables can also be classified as dependent (sometimes called response variables) and independent (sometimes called control variables).

Dependent variables are generally the primary effect or outcome that you are studying.

Independent variables generally influence the primary effect or outcome.

Thus if you are studying how well people cope with stressful social situations, you might use a standardized instrument to measure tolerance to social stress. This instrument would measure the **dependent variable.** If you think serotonin levels influence the ability to tolerate social stress, a **independent variable** might be serum levels of serotonin. Alternatively, since certain blonde-haired and blue-eyed populations tend to have depressed serotonin levels, you might use eye color and hair color as dependent variables.
In this example, we have theorized that the independent variable (serotonin level) influences or controls the dependent variable (tolerance to social stress). Hence the terms *control* and *response* variables are sometimes more intuitive than *independent* and *dependent* variables.
Implicit in this selection of variables is the theory that serotonin levels influence one's tolerance to social stress. In most situations the selection of variables is intimately tied to an underlying theory. This theory is the foundation for the research questions.

A different way in which variables can be classified is by what they measure:

- **Attribute variables** refer to a characteristic of the subject.
- **Quantitative variables** refer to those attributes for which there is a natural numerical measurement.

Examples of attribute variables might be hair color, religious affiliation, political affiliation or gender.
Examples of quantitative variables might be class rank, GRE score or annual income. Quantitative variables fall into three categories:

- **Ordinal** where measures are larger or smaller but magnitude is not specified (class rank is an example).
- **Interval** where a magnitude is specified but there is no natural zero (most standardized psychometric instruments are interval measures).
- **Ratio** where there are both magnitudes and a true zero (a measurement such as annual income is an example of a ratio variable).

A.4. The Model.

At the most fundamental level, the **model** describes the manner in which the variables interact with each other and the context in which that interaction occurs.

Thus when you select your variables, some of which will be dependent and some of which will be independent, you have already described at least in part how the variables interact. The context for the interaction includes the population you are studying and a reasoned explanation of why the variables might interact in the theorized way. This simple model is often sufficient for social science research projects.

Models can be made more complex by adding more detail or structure to the theorized interaction. For example, in a project that studied **income** we might include **annual income** as the dependent variable and **gender** as an independent variable. This selection of variables theorizes that there is relationship between gender and income without specifying the particulars. A more detailed model might theorize that income is lower for females than for males. This provides more detail to the relationship between the two variables.
Sometimes your model can be even more specific. Some of our statistical methods (regression) will let us actually decide whether or not the independent variable can reliably predict the value of the dependent variable. In fact, statistical regression would actually deduce a mathematical formula relating income and gender (where gender were coded as "0" for males and "1" for females, for example). The formula would include an error estimate as well, to give the researcher an idea of the strength of the relationship.

Of course, the **context** of the research will determine when a variable is independent and when it is dependent.
For example, if one were studying voting patterns, then the dependent variable might be **party affiliation** while dependent variables might include both **income** and **gender**. In this example, income is no longer the dependent variable but instead, in this new context, an independent variable.

When defining variables the researcher should keep in mind in mind that each individual has a unique perspective. Diverse research teams can assist bringing varied perspectives to the process, reducing the liklihood that questions will be mis-interpreted.

"Pi what squared? Long John, you should be able to get this."

A.5. Research Objectives.

Your research objective describes what you propose to learn by doing your project. Often research objectives are expressed as hypotheses.

> A **hypothesis** is a provisional idea whose merit is to be evaluated.

A well-formed hypothesis should be **falsifiable**. In order for a hypothesis to be falsifiable

- there must be certain explicit and observable predictions that can be deduced from the hypothesis; and
- it must be possible to gather data, usually through observations, to see whether or not the predictions are correct.

Thus a goal of a research project is often to gather evidence about the predictions of a model. In our prior example on serotonin levels and tolerance for social stress, we hypothesized that changes in serotonin levels would change tolerance for social stress. This is falsifiable since this conjecture meets the two conditions above.
A more detailed conjecture, such as depressed serotonin levels depress tolerance to social stress, is likewise falsifiable.
We are then in a position to gather data and see if the basic predictions of the hypothesis are borne out. If they are, then the hypothesis still is not necessarily proved. It is still provisional since we have failed to disprove it, which is different from proving it! Most hypotheses will go through many years of research and testing before being generally accepted by the scientific community as proved.

Even widely accepted hypotheses can be disproved by the appearance of new data. A striking example was the sudden appearance in 1938 of the coelacanth, a species of fish thought to have been extinct for over 65 million years. The hypothesis that the coelacanth was extinct was falsifiable and, since there were no recent fossils and no known living specimens, the consensus was that the extinction hypothesis was true. The appearance of the fish in the ocean off of South Africa, however, provided evidence proving the hypothesis false.

The lesson of the coelacanth is that scientific knowledge is implicitly provisional. There is always the potential that new data can either refine or overturn the existing scientific consensus.

A.6. Define Measures for the Variables.

At first glance it may appear that the definition of your variables and how you will measure them are the same step. Sometimes this is true and sometimes not.

For example, your variable might be **annual income**. At first glance this would appear to be a quantitative variable of the ratio type. This even has a well-defined measurement provided by everyone's income tax filings.

However, subjects will likely be quite reluctant to reveal to even the most trusted researcher the exact dollar amount of their annual income. This reluctance would greatly complicate data collection and could introduce considerable bias in the study.

A better approach might be to choose to measure this as an ordinal rather than ratio variable by asking subjects "which statement best describes your annual income" and giving them a set of responses:

- $\leq$ *$0.00*
- *$0.00 < $20,000 annually*
- *$20,000.01 < $40,000 annually*
- *$40,000.01 < $60,000 annually*
- *$60,000.01 < $80,000 annually*
- *$80,000.01 or more annually*

Subjects are likely to be more comfortable with this approach. Thus the researcher will most likely get more willing and truthful respondents than with the more intrusive question "what did you report to the IRS last year for your annual income."

A.7. Research Strategy.

The research strategy plots out the specific ways in which data is gathered. The strategy includes for example

- *how the researcher will interact with subjects, both before and after gathering the data;*
- *protecting the rights of the subjects (ethical conduct);*
- *the method by which the researcher will gather data;*
- *whether data will be gathered in an experiment, through observation or by survey.*

One fundamental question to resolve is whether or not the researcher will **observe behavior**, **interview subjects**, or do both. Many research projects will both conduct interviews and observe behaviors. In this case special care must be taken to avoid bias!

Sometimes the act of observing can influence behavior. For example, in conducting a research project to see if people obey the stop signs on campus, would it be a good idea to stand next to the stop sign with a clip board to gather your data?
Some things to consider for observational studies:

- *Will the researcher be a participant in what is being observed?*
- *Will the subjects know that they are being observed?*
- *Will the observations be in a naturally occuring setting or in a contrived setting?*
- *Will a checklist or other structured tool be used to record the observations?*
- *Will observations be direct or indirect (for example, using video or audio recordings)?*

Clearly there are many ethical issues to consider in observational studies since it is possible that the subjects will not have had an opportunity for informed consent.

In your research projects, you may not use minor children in any way without first having the approval of the Institutional Review Board. IRB approval generally takes a minimum of 6-8 weeks to obtain.

Any research involving groups with limited capacity for informed consent (examples would be prisoners or the mentally impaired) must have prior IRB approval before any contact with subjects is undertaken. In all cases care should be taken to protect the confidentiality of respondents

"On the Internet, nobody knows you're a dog."

If the researcher chooses to interact with subjects there are many different ways to accomplish this. Qualitative researchers will generally gather data from such sources as focus groups, case studies, in-depth interviews or expert panels. Quantitative researchers will rely on formal interview techniques, either in a personal interview, in a telephone interview or with a written survey.
If the researcher interacts with the subjects, then special care should be taken to avoid **bias**. Many researchers will carefully script all interactions with subjects, including those interactions prior to administering the survey instrument, to assure that unplanned interactions do not influence subject responses.
Finally, some studies are **cross-sectional** and some are **longitudinal**. A cross-sectional studies gathers point-in-time data while a longitudinal study follows the same group of studies for an extended period of time. The Framingham Heart Study is a famous example of a long-running longitudinal study.

A.8. Research Instruments.

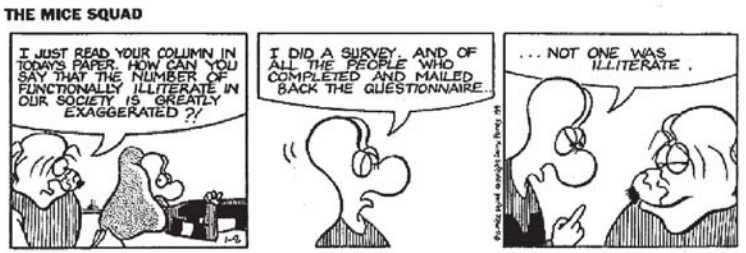
Your research instrument is the tool you use to record your observations. In an observational study this might be a checklist or other structured way of standardizing the observations. In a survey, instrument is the list of questions that you ask.
The researcher should be sure that the instrument measures every variable required for the study. The instrument should not measure variables not needed for the study.

When a researcher does a survey there is necessarily interaction with the subjects. Care must be taken to avoid the instruction of bias in this interaction. When writing survey questions the researcher should

- *Be sure that the survey measures every variable required by your design.*
- *Keep each question focused on one variable (avoid double-barreled questions).*
- *Keep your questions simple. Remember, if a question can be misinterpreted, it will be!*
- *Avoid vague questions.*
- *Avoid leading questions. Sometimes even question order can influence responses!*
- *Make sure that the respondent has enough information to answer.*
- *Be sure that choices are mutually exclusive and collectively exhaustive.*
- *Minimize the number of open-ended questions.*

A.9. Sample Design.

Read the section in the **Study Guide** on this material.

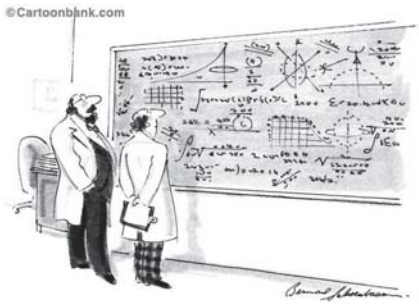


A.10. Budget.

At a minimum your budget should include consideration of

- *Finances*
- *Personnel*
- *Access to subjects*
- *Time constraints*

If your resources are insufficient in any of these critical areas you will need to redesign your project.



"Oh, if only it were so simple."