Chapter 1: Introduction to Behavioral Research on the Internet

This chapter reviews concepts that are essential to research. It cannot replace a good textbook on research methods, but the chapter gives an overview of the key ideas needed to understand research, especially for the examples in this book. It also reviews ways in which Web research differs from laboratory research.

A. The Purposes of Research

The purpose of research is to answer questions. One type of research involves searching for articles and books in the library to find out what other people have discovered or theorized. The other type of research involves making a new investigation to find answers by making controlled observations and measurements. The type of knowledge that comes from gathering evidence is known as *empirical* knowledge. The term *empirical* is distinguished from knowledge based on *theory* or *authority*.

What has been published in the past is not always correct. For example, Aristotle (384-322 BC) theorized that men have more teeth than women. His conclusion was based on the premise that teeth are all about the same size, but men have bigger mouths, and both men and women have mouths full of teeth; therefore, men have more teeth. This theory survived by authority, unquestioned for almost nineteen hundred years, until Vesalius (1514-1564) bothered to count. Vesalius observed that men and women have the same number of teeth. Theories that have been published should not always be accepted on faith or authority; they must be tested for their accuracy. Science progresses by testing theories and proposing new ones to replace those that have been proven wrong.

Psychology is the *science* of the behavior of people and other living organisms. A *science* is the study of alternative explanations. As a study of alternative explanations, it must compare rival or alternative ways to explain the same phenomena.

These ideas can be illustrated by the story of Kluegerhans ("Clever Hans"), the horse who could answer questions by tapping with his hoof. This horse could answer mathematics problems such as "What is 2 + 1?" Kluegerhans would tap out three taps, sigh, and put his hoof down, to the amazement and delight of the audience. Even more amazing was that the horse could answer questions posed in different languages and could even answer factual questions such as "How many days are in a week?"

Two alternative explanations were proposed to explain the behavior of this horse. The first explanation was that Kluegerhans was *sehr klug* (very clever) indeed, that he understood the questions and knew the answers. The rival explanation was that it was a trick, and the horse was getting information from people present.

To test between these two theories, experiments were conducted. In one experiment, the faces of the people in the audience were hidden from Kluegerhans, and it was found that the horse could not do his tricks unless he could see the faces of the audience. In another test, a card that showed 4 spots was shown to the horse, and by a slight-of-hand trick, a different card, with 3 spots, was shown to the audience. Kluegerhans tapped out 3 taps, matching what was shown to the audience, not what he was shown. Apparently, the horse was getting information from the faces in the audience.

The story of Kluegerhans shows the scientific method in action. Psychology is the study of alternative explanations of behavior. The behavior to be explained was the behavior of the horse: how was the horse able to answer questions? Second, the two

alternative explanations were theories of how the horse got right answers. One of these theories predicted that he should continue to be right, even if the audience is hidden or if they were shown a different card, because he knew the right answers. The other theory was that the horse simply tapped until the audience gave a clue, by blinking and raising their eyes, that the right answer had been reached. The experiments were designed to test between these rival explanations.

B. Philosophical Criteria for Explanation

Explanations are theories. In psychology, they are theories of behavior. An explanation satisfies five criteria. An explanation is deductive, meaningful, predictive, causal, and general. Each of these criteria deserves some discussion.

The first criterion of a theory is that it is deductive. Given the explanation, one can *deduce* the event to be explained from the explanation. For example, suppose we wanted to explain the behavior of Kluegerhans, and someone said, "The horse got the answers right because he was a horse." The sentence that Kluegerhans was a horse is correct, but that fact does not explain the behavior.

It would be deductive to say, Kluegerhans is a horse; all horses can answer questions correctly; therefore, Kluegerhans can answer questions correctly. That would be deductive, but just plain false, because not all horses can do those tricks.

Theories are to data (observations), as premises are to conclusions. True premises and logic lead to true conclusions. However, a correct deduction with false premises can lead to a true conclusion. For that reason, one cannot use the truth of a conclusion to argue for the truth of the premises. For example,

Everything made of Cyanide is good to eat.

Bread is made of Cyanide.

Therefore, Bread is good to eat.

The conclusion is true, but the premises are false. A person can not "prove" the theory true by eating bread, showing that it is good to eat. To *test* a theory, one makes observations that might show the theory is *not* true. However, if the conclusion is false, then something is wrong with the premises. Thus, a theory can be disproved by experiment, but not proved.

Second, an explanation is *meaningful*. The *meaning* of a sentence is equivalent to the set of testable implications of the sentence. If a sentence has no implications, then it is meaningless. For example, suppose someone suggested the following theory of behavior:

"Everything is caused by the action of invisible, undetectable, and logically unverifiable brownies. These brownies cause everything. If they did not exist, nothing could happen. Because things happen, we know the brownies exist. They compete with each other to cause events in the world, and the more dominant brownie always wins. These brownies cannot be tested, observed, or measured in any way. But they exist, they cause everything, and they explain everything."

The problem with this so-called "theory" is that it is meaningless. If the Brownies are logically unverifiable, then it is a contradiction to assert that one could obtain test their existence. There is no experiment, no test, no action we can take to find out if the theory is true or not. Such ideas are meaningless, and one should not waste one's time arguing about them. It is sometimes surprising how much energy can be spent arguing over matters that cannot be put, even in principle, to empirical test.

The third criterion of a theory is that it is *predictive*. If the explanation were known in advance, the behavior would have been predicted, in principle. Given a good theory of the solar system, one should be able to predict eclipses of the sun and moon, and the relative positions of the planets in the sky on any date in the future. If a "theory" could not predict events (e.g., eclipses) until after they happen, then it is called *post hoc*. A good theory, if known in advance, would allow one to predict what will happen next.

However, prediction is not enough. Suppose someone asked, "Why were there 10 murders in a certain town last year? Suppose another person answered, because that town has 1000 telephone poles, and the number of murders in a town can be predicted from the equation $Y = .01*X$, where X is the number of telephones, and Y is the number of murders. The above system is meaningful (we can test if it is true for all towns and we can test if $X = 1000$), it is deductive (we can deduce from X what Y should be), and it is predictive (knowing $X = 1000$, we predict $Y = 10$). However, the theory does *not* tell us what would happen if someone chopped down those telephone polls. It is not causal.

The fourth criterion of a theory is that it is *causal*. In principle, the explanation of a behavior tells one how to *control*, or change the behavior in question. The idea of *in principle* means that we may not be able to make the change in practice. For example, in Astronomy, there are theories of what would happen if two stars were made to combine their masses, but no one is yet able to actually cause that to happen. The theories are still causal because they dictate what would happen if one could combine stars.

Causation must be distinguished from correlation. The number of telephone polls in a city is *correlated* with the number of murders in a city. One can use the number of telephone polls to predict the number of murders, or vice versa. But correlation does not

allow one to predict what would happen if the number of telephone polls or murders was made to change. Causation, in contrast, allows one to predict what would happen if we chopped down the telephone polls and buried the cables underground. Chopping down the telephone polls in some cities but not others would involve an *experiment*. Experiments allow one to predict the effect of changes; they allow one to test causal hypotheses. More will be said about causation and correlation in Chapter 10.

C. Causal Experiments

The purpose of an experiment is to test a causal *hypothesis*. An *hypothesis* is a conjecture that has not been proven or established. A causal hypothesis is to be distinguished from the *null hypothesis*, which is that there is no causal relation. The key idea of an experiment is control. The experimenter controls, or causes the suspected cause to happen and examines if the suspected effect occurs.

To illustrate a causal experiment, consider the hypothesis that penicillin would cause a reduction in the probability of death among people with Strep infections. Strep infections are diagnosed by high fever, sore throat, and white streaks in the throat that (when sampled and examined under the microscope) contain Streptococcus bacteria. This hypothesis has already been tested so you may already be convinced. However, please imagine yourself in 1920, before the safety and effectiveness of penicillin were known.

In an example of the classic *double-blind experiment*, people with Strep throats are randomly assigned to one of two conditions. By *random assignment* is meant that a coin is tossed (or some other random mechanism is used) to decide which condition each person will receive. In one condition (the *treatment group*), patients receive pills with

penicillin. In the other condition, called the *control group*, patients receive pills that look identical, but contain no penicillin. The pills that contain only the inert ingredients (starch and flavors) are known as *placebos*. A *placebo* is an inert treatment used with the control group. Without the control group, we do not know how many people would have lived if the disease went untreated. There would be no way to evaluate the treatment.

The experiment is called a *double-blind* study because the patients who receive the medicine and the doctors who administer the medicine are both "blind" with respect to whether the pills contain penicillin or just placebo. The reason to keep the patients blind, is that when people think they have received a powerful medicine, they often get better from the Voodoo effect. In Voodoo, one can make another person sick by suggesting that pins stuck in a doll will cause pain and injury to the person who is represented by the doll. Similarly, placebos can help a person get better; the effect can be as powerful as some medicines. Many new medicines have side effects that can make a person sick or even dead; therefore, a placebo is often superior to a pill because it is less dangerous. In evaluating a new medicine, it is important to show that it is more effective than a placebo.

The reason to keep doctors "blind" is to keep them from switching patients from the experimental and control groups. Often a doctor will try to "help" make the new medicine look good. Young, strong, and otherwise healthy people are moved to the treatment group, and old, infirm, and otherwise unhealthy people are moved to the placebo group. That might make a harmful medicine look good. But that would deceive the scientists and future patients, who really want to know if the medicine is beneficial. By trying to "help" doctors (and even experimenters) can ruin experiments. If doctors

are not "blind," they may reveal (by subtle cues) to the patient that they do not expect the patient to live, and such attitudes can become self-fulfilling prophesies.

In some studies, there is also a third person who should be "blind" with respect to the treatment; that is the judge who rates the success of the treatment. In studies of acne cream, for example, one might use judges to rate how "pimply" the faces of patients appear. It is important to keep the judge from knowing if the patient received the medicine or the placebo. Such studies in which the patient, the doctor, and the judge are "blind" with respect to the treatments are called *triple blind experiments*. In order to keep these people "blind" bottles of medicine are coded with numbers that the experimenter has recorded to keep track of which bottles have medicine, and which have placebo.

Suppose in the penicillin study, there were 200 patients with Strep throats. Suppose 100 were assigned to each group. Suppose 80 of the treatment group lived to the end of the year, and 20 in the control group lived to the end of the year. What can we conclude? Apparently, more people lived in the placebo condition, but this might have happened by chance.

Independent Variable: Random Assignment to	Dependent Variable	
	Dead	Alive
Placebo (n = 100)	80	20
Penicillin (n = 100)	20	80

If one had 200 cards, of which 100 said "dead" and 100 said "alive," there is a chance that if they were mixed randomly and dealt into two piles, one pile might have 80

alive and the other might have 20. How do we know if the treatment was effective, if the result might have happened by chance?

The answer is that statistics allows us to calculate the probability that such a difference might have happened by chance, given the null hypothesis that the treatment had no effect. In this case, the probability of getting such an extreme result, given the null hypothesis, is less than one chance in twenty. When the probability is small that the result occurred by chance, one rejects the null hypothesis. Therefore, we reject the hypothesis that the data occurred by chance in favor of the hypothesis that the result was caused by the difference in treatments. In this case, the result indicates that people who received penicillin are *significantly* less likely to die than those who did not. The term *significant* is used to denote that the result is extremely unlikely by chance alone, according to the null hypothesis.

When the statistic is *not* significant, it does *not* prove that the null hypothesis is true. For example, a researcher might have done the penicillin study with 10 patients, and fount that 4 of the 5 treated survived and only 1 of 5 who received placebos survived. That is the same proportion as in the larger study; however, this result is not unusual enough to warrant rejection of the null hypothesis. Failure to reject the null hypothesis does *not* show that the null hypothesis is true.

The situation is like looking for a key in a large wheat field. If one finds the key, one can reject the null hypothesis that there was no key. On the other hand, if one looks and does not find the key, it does not prove that there was no key, it just means one did not find it. Maybe the key is still there, and might be found by a more thorough search. Thus, there are really three possible conclusions of a study: The null hypothesis may be

false, it may be true, and the data may yield no answer. Non-significance means no conclusion.

The *power* of an experiment refers to the probability of finding significant results if the null hypothesis is false. The more powerful our search, the more likely we will find the key if it is there. Power increases as measurements become more precise and as the sample size is increased. As noted below, Web research is usually more powerful than lab research.

D. Generality

The fifth criterion of a theory is that it is *general*. By *general* is meant that one can deduce from the theory not only the behavior that was to be explained, but also more implications for new experiments that give the explanation *predictive power*. The greater the number and variety of implications, the greater the generality of the theory.

People were extremely impressed when Newton (1642-1727) showed that a few simple ideas plus some mathematics could be used to derive implications not only for the motions of objects on earth, but also for the motions of the planets. The number and variety of different experimental implications were extremely large, and the few simple premises could be used to predict the results of thousands of experiments involving motions and collisions of objects. Because the theory made so many interesting predictions that could be tested by a variety of experiments, the theory was recognized as having great generality.

E. Experimental Designs

The classic, two groups design is not the only way to do research. In fact, in psychology, certain types of between-groups studies can lead to strange conclusions such

as that the number 9 seems "bigger" as a number to people than does the number 221 (See Chapter 9).

The penicillin study is termed *between-subjects* or between-groups because the experimental variable is different for different people—each person gets only one treatment, either placebo or penicillin.

Why is random assignment to groups important? Why not let some people volunteer to take the new medicine and let the others be in the control group? The answer is that those people who volunteer might be those who are younger and more optimistic. If so, then they are more likely to live anyway, even if they did not receive the drug. On the other hand, those who volunteer might be those who are the most sick and desperate; perhaps these would be more likely to die anyway. Therefore, we must not allow such variables to be *confounded* with the treatment. The term *confounded variables* refers to variables that might confuse or confound our attempts to learn if the medicine works.

The variable that is suspected as the cause is made *independent* of all other variables by random assignment to conditions. The *independent variable* is the variable that the experimenter manipulates. The suspected effect, or *dependent variable*, is the variable that the experimenter measures to assess the effect of the treatment. Sometimes people use the term "independent variable" for a variable that they *believe* is the cause, even though the variable is not independent of other variables. Such misuse of terms is called *deception*.

A classic study that disentangled confounded variables was the study by the royal commission appointed to investigate if animal magnetism was a real phenomenon. Mesmer (1734-1815) created a great sensation in France by putting people into trances by supposedly using animal magnetism. Mesmer or his assistants could mystically "magnetize" or

mesmerize an object, and a person who touched the object would fall into a trance. The commission, which included the American scientist and publisher, Ben Franklin (1706-1790), realized that the *knowledge* that the object had been mesmerized was confounded with the actual procedure of mesmerizing it. They created a within-subjects, factorial design in which a tree either was or was not mesmerized, and each person was told either that the tree was mesmerized or not. This experiment revealed that trances only occurred when people were told that the tree had been mesmerized, and that there was no effect of actually "magnetizing" the tree. By teasing out the beliefs of the person from the actual procedures of magnetizing an object, the commission concluded that the phenomenon was psychological and not due to any new magnetic or electric force.

It is important to emphasize that a variable is something that varies. The independent variable in the penicillin study is the *difference* between placebo and penicillin. The dependent variable in that study is the *difference* between being dead and alive. The independent variables in the study of animal magnetism were belief that the tree was mesmerized or not, and whether the tree was mesmerized or not. The dependent variable was falling into a trance or not.

Variables must therefore have at least two levels, but they can take on more than two levels. For example, in a study of a disease that does not kill, one might measure the number of days each patient spends at home (away from work) as the dependent variable. The independent variable can also have a number of levels. For example, one might study 4 different doses of a new medicine, with a fifth level as the placebo. The purpose of such research would be to find an optimal dose that is both effective and safe from side-effects. If the dose is too small, the medicine may not be effective; if the dose is too

large, it may be toxic. Before putting the drug on the market, it is important to find the optimum level of drug to accomplish its purpose without harming the patient.

Experiments can include more than one dependent variable. For example, in a drug study of the effectiveness of a cold remedy, dependent variables might include number of days that the patient feels "sick", the number of days with fever, the temperature of the highest "spike" fever, the number of days with body pains, etc.

Experiments can also include more than one independent variable. Factorial designs are experiments that are designed to investigate how two or more independent variables combine. For example, consider an experiment on the effects of two antibiotics for patients in hospitals with "fever of unknown origin." Such fevers may be due to nosocomial infections—those are infections that are given to patients by nurses or doctors who do not wash their hands or change gloves between patients. Failure to follow these precautions spreads disease within modern hospitals from one patient to another. More people are murdered in hospitals each year this way by germs than are killed by bullets and traffic accidents combined.

Suppose there are two antibiotics that are effective against different strains of bacteria. Suppose each antibiotic would be shown to be effective if tested against placebo in a double-blind, study with two groups. However, suppose the combination of antibiotics is poisonous, causing damage to the kidneys. The only way to find this out would be in a factorial design with (at least) four treatment combinations, as shown in Table 1.1. Insert Table 1.1 about here.

Table 1.1. Results of hypothetical Factorial Drug Experiment. Each entry is the percentage of patients (who had "fever of unknown origin" in the hospital), who survive for one year after their stay in the hospital. Note that each drug is effective when tested against a placebo; however, in combination, they are worse than no treatment at all.

	Placebo A	Drug A
Placebo B	50	80
Drug B	70	30

In the factorial design, there are four groups of patients in a double-blind study. Patients are randomly assigned to the four conditions. Every patient gets two bottles of medicine, and is instructed to take one dose from each bottle. The first group receives two placebos. The second group receives Placebo B (made to look like Drug B) *and* Drug A. The third group receives Placebo A (made to resemble Drug A) *and* Drug B. The fourth group receives Drug A *and* Drug B. There are two independent variables in this study, Drug A versus Placebo A and Drug B versus Placebo B.

The first row of the factorial design is a simple test of Drug A (two groups, double blind experiment). The first column of the factorial design is a simple test of Drug B. The factorial design also includes the combination of *both* A and B. If each drug were effective against a different type of bacterium spread in the hospital, one might hope that the results would be additive, in which case the combination would be more effective than either drug alone. However, the hypothetical data show that the combination is worse than no medicine at all. This would be an example of an interaction between two factors. In this case, the drug interaction is harmful. In another type of interaction, two drugs might be more effective in combination than the sum of each one separately.

F. Within-Subjects Research

Some studies are done *within-subjects*. In a within-subjects experiment, each subject receives all of the treatment combinations. For example, one might have a drug study to investigate the effects of two ingredients in a cold tablet—aspirin and caffeine. A factorial design would mean that there are four types of pills to be tested: those that contain two placebos, those that contain aspirin only, caffeine only, or both aspirin and caffeine. Pills might be labeled with labels such as A, B, C, D. Each patient would be

instructed to take Pill A for one cold, Pill B for the second cold, Pill C for the third, and Pill D for the fourth.

Within-subjects designs pose new issues that must be handled. For example, suppose the labels of the drugs has an effect. Maybe people would like a pill labeled "A" more than one labeled "B". Similarly, perhaps the first cold in a flu season is more severe than subsequent colds. To counterbalance the effects of the labels, one might use a Latin-Squares design, such as is illustrated in Table 1.2. Insert Table 1.2. about here.

Notice that in the Latin-Squares design, each label is equally often applied to each treatment, and each treatment gets each of the labels. Four groups would receive the drug combinations, but each group would be in one of the rows of the Latin Square. Thus, in the first group, the Placebo is labeled A, but in the second row, the Placebo is labeled D, then C, then B.

Table 1.2. Latin Square design for Pill Labels. In this design, the column means represent the treatment effects, and differences among the rows represent effects of labeling.

Group	Placebo	Ingredient 1	Ingredient 2	Both ingredients
1	A	B	C	D
2	D	A	B	C
3	C	D	A	B
4	B	C	D	A

The order of taking the medicines can also be counterbalanced. One group would be instructed to take pill A, followed by B, then C, then D. Another group might be instructed to take pill D then C then B then A. Thus, the treatment order can be counterbalanced, either by a Latin-Square, which would be created within each row of Table 1.2, or in a complete counterbalancing of all 24 possible orders.

Counterbalancing of such factors does two things. First, it averages out and disentangles the effects of unwanted phenomena, such as the effects of the labels of the medicines. Second, it allows the investigator to study these effects, to see if they are substantial in magnitude. If the label of the pill does in fact have a large effect on how people respond to the medicine, then this information itself is of value.

Notice that when a within-subjects experiment is properly counterbalanced, one can always look at the first treatment as a between-subjects design. Therefore, a counterbalanced, within-subjects study contains a between-subjects study. Within-subjects experiments are more powerful than between-subjects experiments, and they have other advantages in behavioral research that will be explained in Chapter 9. More information about within-subjects, factorial experiments will be presented in Chapters 11—16.

Not all research uses causal experiments. Some research is concerned with prediction. There can be value in being able to predict which parole candidate will commit crimes if given parole, which candidate for law school will flunk out in the first year, or which person will quit a job after receiving training. Prediction research is based on correlational methods. The researcher collects several dependent variables and

computes correlations among them to see if some variables are related to others. This type of research will be treated in Chapter 10.

G. Web Research

Since 1995, it has become possible to conduct research by a new method that has several advantages over traditional laboratory research. This new method uses the Internet as a medium for conducting behavioral research. The Internet allows one to collect data not only in laboratories with Internet-connected computers, but also to collect large quantities of high-quality data from people all over the world. At the same time, there are special considerations and potential problems that require additional skills in the design and execution of Internet research.

At the present time, few researchers are trained in the techniques that must be mastered to conduct meaningful research on the Web. The purpose of this book is to provide the background needed to conduct this type of research. The skills that will be covered in this book will be of lasting value, not only to those who plan to do graduate research in the behavioral sciences, but also to students who plan to enter the work force, where expertise in the Internet has become a valued asset.

H. About the Web

In the last decade, a new protocol for the exchange of information between computers was introduced to the Internet. The *Internet* refers to the network of computers that are connected to each other and exchange information by email and other protocols. The new protocol introduced in 1990 was *http*, which stands for *Hypertext Transfer Protocol*. By 1993, there were about 100 computers communicating by this

new protocol, and this network of computers and the information they contained became known as the *World Wide Web*. (WWW), or simply, the *Web*.

The Web pages sent via this new protocol contained commands (*tags*) in a new language known as Hypertext Markup Language (HTML), which displayed text, graphics, and other multimedia, and which allowed one to link one portion of a document to other information on the Web. These HTML files became known as Web *pages*. When Web *browsers*, programs that use a graphical interface to load and display Web pages, were introduced, the Web grew at an astonishing rate.

As computers and their software became more powerful, less expensive, and easier to operate, more and more institutions and individuals became attached to the Web. Web *servers*, computers that "serve" files in response to requests from remote browsers, were now affordable to small organizations, research laboratories, and individuals. At the same time, the new languages of Java and JavaScript were introduced. These languages allowed programs delivered with Web pages to remotely control distant browsers, even though those remote browsers might be running with different systems on different platforms (e.g., PC, Mac, etc.), and they might be on the other side of the world.

In 1995, the new standard of HTML supported Web forms, which allow one to receive data from a person using a Web browser. At this time, a few behavioral researchers began to collect data via the Web. The early "pioneers" of Web research soon found that it was possible to collect large quantities of high-quality data in this way. A number of these investigators have shared their experiences and contributed much good advice for other professionals in a book edited by Birnbaum (in press). Musch and Reips (in press) surveyed the pioneers of Web experimentation, and found that they were

quite pleased with their Web experiments and planned to conduct future research that way.

The number of Web studies listed in the American Psychological Society doubled from 1998 to 1999. It is reasonable to predict continued growth in this type of research, since it has many advantages over the most common type of laboratory research done in the behavioral sciences (Schmidt, 1997; Reips, in press). Studies that have compared Web and lab research on the same topic have found that these two research methods lead to the same conclusions (Birnbaum, in press-b; Buchanan, in press; Buchanan & Smith, 1999; Krantz, Ballard, & Scher, 1997; Krantz & Dalal, in press; Pasveer & Ellard, 1998; Pettit, 1999; Stanton, 1998).

The typical study in psychology is conducted using paper and pencil methods. The data are collected via questionnaires from "subjects," usually college students recruited from a "subject" pool of people who will receive credit toward an assignment in lower division psychology. A research assistant, in a laboratory collects the data at a prearranged time. The assistant then codes the data and enters them in the computer for statistical analysis. Additional time is required to verify that the data have been properly coded and entered, and to fix any errors. In contrast, a typical Web experiment allows the participant to complete the materials on-line, the data are coded by computer, and saved to a data file for immediate analysis. The time required to conduct a study may be reduced by a factor of ten or more (Birnbaum, submitted).

I. Comparisons of Web and Lab Research

One can compare Web and laboratory research with respect to the dimensions listed in Table 1.3. Laboratory research is typically conducted with students in

psychology courses. In the past, psychology enrolled an even mix of males and females. At the present time, however, the "subject pool" in psychology is predominantly female. The Web was once considered predominantly male; however, participants in Web experiments are more nearly equal in sex ratios than the subject pool, and recent studies show that more females than males participate in on-line psychology studies (Birnbaum, in press-b).

Insert Table 1.3 about here.

College students are very homogeneous in education: they have all graduated high school and have not yet graduated college. On the Web, participants are more heterogeneous. There are people who dropped out of school in the eighth grade, there are college graduates, and there are a good number who have advanced degrees. The comparison on age is similar. Because samples recruited from the Web are so heterogeneous on such demographic characteristics, one can divide the sample on these variables to examine if the data support the same conclusions within each demographic group.

Web samples can also be recruited by techniques that are designed to reach specialized, rare populations. For example, identical twins are fairly rare in the general populations, but could be easily recruited via the Web from on-line groups such as Mothers of Twins clubs. Similarly, transvestites constitute perhaps 1 or 2 percent of the general population, but can be contacted via social clubs that have Web sites.

Lab research must be conducted in a special place at a present time. An assistant must unlock the door, greet the participants, and conduct the test or experiment. Web research collects data around the clock and around the world. Participants typically come

on line at times convenient to them from computers at home, school, or work. For these reasons, it is possible to collect large samples with high power on the Web.

In many universities in which many professors and students have active research programs, it is not possible to test large numbers of subjects. In addition to the ethics review, which ensures that the benefits of the research outweigh any risks and that participants are treated with respect, many universities also have an allocation of subject-hours from the limited "subject pool". In contrast, once a Web project is approved by the ethics committee, there are currently no limits on the number of people that can be tested. In some universities, an investigator would be lucky to obtain 200 subject-hours per year from the subject pool. On the Web in 1999, it was quite reasonable to test 6,000 people per year. Thus, the amount of data that might be collected in a year in the lab can be obtained in weeks on the Web.

Because the Web is World Wide, it is possible to test people in other countries from other cultures. Without the expense of travel, it is possible therefore to do cross-cultural research. To do such research in the lab, one would need to either travel, or at least to make contact with colleagues in different countries who can collaborate on the cross-cultural project.

Of course, the Web is not used by "primitive" people (people who do not use computers). Therefore, anthropological research will still require a person on the scene. Nevertheless, an anthropological research armed with a laptop can send data to home base instantly via the Web.

Web research cannot be used to investigate issues that require the experimenter to have "hands on" the subject. It is not possible, for example, to inject drugs, to measure

EEGs (Electro-encephalo Grams), take PET scans, image X-rays, do surgery, or other such manipulations and measures that require one to have “hands on” access to the subject.

On the other hand, some Web research may benefit from there being no experimenter to introduce “experimenter bias” to the results. In some studies, it has been found that experimenters seem to bias the subjects to produce certain results. They may unintentionally give subtle cues that reinforce certain behavior by participants. They may do this by obvious methods, such as giving instructions that “help” the experiment to work, or they may even alter data as they are reported by the subject. With lab research, experimenter effects can cause investigators at different universities to find contradictory results that may require a great effort in the lab to resolve.

Web studies have a situation that is familiar to users of the Internet (the browser interface), a situation that can be easily described and replicated. Thus, with Web research, it is possible to standardize the situation, allowing exact replication. Web studies can be made public to other scientists, making the process of research more open. When everyone can do the same experiment, different investigators can produce the same results.

The benefits of doing research on the Web justify the efforts required to learn how to do this new style of research.

Table 1.3. Comparisons of Traditional Laboratory Research against Web-based research.

Aspect	Lab Research	Web Research
Sample Characteristics	Not random: College students	Not random: depends on recruitment.
Education	Homogeneous	Better educated; more diversity
Age	Most 18-23 years	Heterogeneous; older
Occupation	Temp jobs, minimal	Heterogeneous; careers
Gender	Mostly female	More equal sex ratio
Specialized Samples	Impractical	Recruit via Internet
Large Samples	Impractical or costly	Easy to accomplish
Cross-cultural research	Difficult or impossible	Fairly easy to do
Equipment, space needs	Considerable	Minimal
Data coding, entry	Expensive, time-consuming	By computer
Lab assistants	Necessary	Not required
Experimenter effects	Relevant	Avoided, uniformity
Variety of Experiments	Equipment, Drugs, Surgery, Scans, etc. possible	Many I.V. and D.V. not possible.
Data quality	Good	Higher by some comparisons.
Control	Highly controlled; depends on experimenter present.	Less controlled conditions; depends on programming.
Interface	Unfamiliar	Familiar
Drop-outs: Between-Ss.	A problem	Bigger problem
Motivation	Credit to class assignment, pay or incentives	Volunteers, interests, incentives offered
Ethical Review	Unit IRB	Unit IRB, some new issues
Multiple Submissions	Rarely considered	A concern; handled by data checking.

J. Summary

This chapter reviewed the goals and terminology of basic research. The purpose of research is to answer questions, in particular questions about explanations of behavior. Behavioral scientists use experiments to investigate causation. The concepts of independent variable, dependent variable, and random assignment to conditions were defined and illustrated with examples. Differences between Web research and lab research were considered. It was noted that Web research can more easily achieve large samples, which allow powerful tests of the null hypothesis.

K. Exercises

1. What are the philosophical criteria for an explanation? Why is each criterion needed?
2. Why is random assignment to conditions needed in a study to investigate the effects of a new drug?
3. Why is a control group needed?
4. What is a placebo?
5. Define the following terms, independent variable, dependent variable, treatment group, control group.
6. Why do within-subjects designs require counterbalancing?
7. What are the chief differences between Web and lab research?
8. Distinguish causation from correlation.
9. In the experiment to test if Kluegerhans could answer questions if the faces of the audience were covered, what were the independent and dependent variables? Was it a within-subjects or between-subjects design?