

## Section III: Experiments

IN AN OBSERVATIONAL STUDY, RESEARCHERS MAKE OBSERVATIONS AND RECORD DATA. As much as possible, the observer tries not to influence what is being observed. In an experiment, researchers deliberately do something and then measure a response. The “participants” in an experiment are called **experimental units**. Experimental units can be people, animals, or objects. When the experimental units are people, they are often referred to as **subjects**. The specific conditions researchers impose on the experimental units are called **treatments**. As experimental units may differ from one another in many important ways, the method of assigning treatments to experimental units is an important concern in the experimental design process.



Let's look at an example. A biologist would like to determine which of two leading brands of weed killer is less likely to harm the broad-leafed plants in a garden at the university. Before spraying near the plants in the garden, the biologist decides to conduct an experiment that will allow her to compare the effects of these two brands of weed killer on broad-leafed pansy plants (one of the varieties in the garden). The biologist obtains 24 individual pansy plants to use in the experiment. In this simple experiment, the *experimental units* are the individual pansy plants and the *treatments* are the two brands of weed killer.

Consider the following two plans for assigning treatments to the pansy plants:

**Plan A:** Choose the 12 healthiest looking pansy plants. Apply brand X weed killer to all 12 of those plants. Apply brand Y weed killer to the remaining 12 pansy plants.

**Plan B:** Choose 12 of the 24 individual pansy plants at random. Apply brand X weed killer to those 12 plants and brand Y weed killer to the remaining 12 plants.

Which plan seems preferable? Let's evaluate what might happen with each of these plans.

Under Plan A, suppose the pansy plants treated with brand Y weed killer have many more dead or dying leaves than the pansy plants treated with brand X. Can the biologist feel confident recommending brand X to the campus gardener as the safer weed killer? Not at all. Since the healthier plants received the brand X treatment and the less healthy plants received the brand Y treatment, it could be that more leaves were dead or dying on the pansy plants treated with brand Y because those plants were less healthy to begin with. We really can't separate the effects of the two brands of weed killer from the effect of the original healthiness of the plants in the two groups. The inability to separate the effects of the treatments from the effects of another variable in a study is known as **confounding**.

With Plan B, individual pansy plants are assigned at random to one of the two weed killer treatments. This **random assignment** helps to ensure that the group of plants treated with brand X and the group of plants treated with brand Y are fairly similar to begin with in terms of all characteristics that might affect the plants' responses to the treatments. If the biologist then observes that the pansy plants treated with brand Y

weed killer have many more dead or dying leaves than the pansy plants treated with brand X, there are two plausible explanations for the observed difference.

First, it is possible that there is no difference in the effects of the two brands of weed killer on pansy plants. Some pansies are heartier than others, and, just by chance, the random assignment placed more of those healthy plants in the group that was treated with brand X. In other words, the observed difference could be simply due to chance.

The second possible explanation is that brand X weed killer actually results in greater harm to pansy plants than brand Y. In that case, we could say the difference in the number of dead or dying leaves between the two groups of pansy plants is a direct result of the brand of weed killer used. Put another way, the difference in brand of weed killer *caused* the difference in the number of dead or dying leaves.

Random assignment of treatments to subjects is an essential component of well-designed experiments. One of the big advantages of such experiments is their ability to help the researcher establish that changes in one variable (like brand of weed killer) cause changes in another variable (like number of dead or dying leaves). Since establishing causation is often a goal of experiments, we find it useful to give names to the two variables mentioned in the previous sentence. We call the variables that the experimenters directly manipulate the **explanatory variables** or **factors** and the variables that measure the subjects' responses to the treatments the **response variables**. The treatments in an experiment correspond to the different possible values of the explanatory variables. For the weed killer experiment above, there is one factor—brand of weed killer—and one response variable—number of dead or dying leaves.

In addition to randomly assigning treatments to experimental units, there are two other important considerations in designing experiments. The first is to **control** for the effects of variables that are not factors in the experiment but that might affect experimental units' responses to the treatments. Some variables can be controlled by trying to keep them at a constant value. For example, the biologist would want to ensure that the plants all receive the same amount of water and are exposed to the same amount of light. If everything is roughly equivalent for the two groups of plants except for the treatments, and we observe a difference in the response variable, then that difference is either a result of the random assignment or is caused by the difference in treatments.

Some variables can't be easily controlled by keeping them at a constant value. One such variable in the weed killer example was the current state of health of the plant. In this case, the random assignment of plants to treatments should help spread the healthy and less healthy plants out in a fairly balanced way between the two groups of pansy plants. Then, any differences in the number of dead or dying leaves that appear should not be a result of differences in initial plant health.

The other important experimental design principle is **replication**. In a nutshell, replication means giving each treatment to enough experimental units so that any difference in the effects of the treatments is likely to be detected. Imagine the biologist treating one pansy plant with brand X weed killer and one pansy plant with brand Y weed killer. If the plant treated with brand Y has more dead or dying leaves, can the biologist conclude that brand X is safer to use on the university's pansy plants? Of course not. Individual pansy plants vary widely in terms of general health and other characteristics that might affect their response to a particular brand of weed killer. With only one experimental unit available for each treatment, the random assignment can't be counted on to produce roughly "equivalent" groups prior to administering the treatments. Any difference we observe in the number of dead or dying leaves on the two pansy plants could simply be due to the difference in the initial health of the plants.

Now imagine the biologist conducting the same weed killer experiment, but with 50 pansy plants receiving each treatment. If the pansies treated with brand Y have a much higher number of dead or dying leaves than the pansies treated with brand X, the biologist should feel much more confident concluding that the difference in treatments caused the observed difference in the response variable.

Let's look at one more example. In the fall of 1982, researchers launched a now famous experiment investigating the effects of aspirin and beta carotene on heart disease and cancer. Over 22,000 healthy male physicians between the ages of 40 and 84 agreed to serve as *subjects* in the experiment. The two *factors* being manipulated by the researchers were whether a person took aspirin regularly and whether a person took beta carotene regularly. Researchers decided to use four treatments: (1) aspirin every other day and beta carotene every other day, (2) aspirin every other day and "fake" beta carotene every other day, (3) "fake" aspirin every other day and beta carotene every other day, and (4) "fake" aspirin every other day and "fake" beta carotene every other day.

The "fake" pills looked, tasted, and smelled like the pills with the active ingredient, but had no active ingredient themselves. (We call such "fake" treatments **placebos**.) Subjects were randomly assigned in roughly equal numbers to the four groups. Several *response variables* were measured in the study, including whether the individual had a heart attack and whether the individual developed cancer. Neither the subjects nor the people measuring the response variable knew who was receiving which treatment. We say this experiment was carried out in a **double-blind** manner. If either the subjects or the people measuring the response variable knows who is receiving which treatment, but the other doesn't, then the experiment is **single-blind**.

An outside group of statisticians that was monitoring the Physicians' Health Study reviewed data from the experiment on a regular basis. To everyone's surprise, the data monitoring board stopped the aspirin part of the experiment several years ahead of schedule. Why? Because there was compelling evidence that the subjects taking aspirin were having far fewer heart attacks than those who were taking placebo aspirin. It

would have been unethical to continue allowing some physicians to take a placebo with clear evidence that aspirin reduced the risk of heart attack.

Even though the Physicians' Health Study was an exceptionally well-designed experiment, it does have some limitations. Researchers decided to use male physicians as subjects because they felt doctors would be more likely to understand the importance of taking the pills every other day for the duration of the study. That may be true, but because only male physicians were used in the study, we cannot generalize the findings of this study to women, or even to all male adults. We can feel pretty confident concluding that taking aspirin regularly *caused* a reduction in heart attack risk. However, the benefits of taking aspirin regularly might be offset by other effects of the drug, such as an increased risk of stroke. In spite of its limitations, the Physicians' Health Study provided a template for other researchers who wanted to design experiments to help answer important questions.

In many published reports of experimental studies, we see conclusions such as “the observed difference in heart attack rates was **statistically significant**.” This tells us that the differences in the response variable between those in different treatment groups cannot reasonably be explained by the chance involved in the random assignment of treatments to subjects. Recall what we said earlier: There are only two possible explanations for the observed differences in an experiment—that they were due to the chance involved in the random assignment or that the difference in treatments caused the difference in the response variable. Saying that the results of a particular experiment are *not* statistically significant means that we can't rule out the possibility that there is no difference in the effects of the treatments, and that the differences in response are simply due to the random assignment.

You may have noticed that in both the examples presented here, the subjects were *not* randomly selected from a larger population. This is usually the case with experiments. It often isn't practical to choose subjects at random from the population of interest. Consider how you would go about randomly selecting 24 pansy plants from the population of *all* pansy plants, for example. Or how researchers might randomly select 22,000 male physicians. As you learned earlier, the lack of random selection limits our ability to generalize results to the population of interest.

However, even if experimental units are not randomly selected, well-designed experiments can give convincing evidence that changes in one variable cause changes in another variable. Establishing causation is much more difficult with observational studies, because researchers cannot hold other variables constant and cannot assign individuals at random to treatment groups. As an example, consider early observational studies that suggested people who smoked were much more likely to get lung cancer than people who didn't smoke. Cigarette company executives argued that *confounding* was at work. They claimed that the kinds of people who smoked were also much more likely to engage in other unhealthy activities—such as drinking, overeating, and failing

to exercise—than people who didn't smoke. It was these other unhealthy behaviors, they said, that led to increased risk of cancer, not smoking cigarettes. After many other observational studies showed the strong connection between smoking and lung cancer, and experiments on animal subjects demonstrated that smoking caused cancerous growths, cigarette company executives finally conceded.

There are only two possible explanations for the observed differences in an experiment — that they were due to the chance involved in the random assignment or that the difference in treatments caused the difference in the response variable.





# Investigation #10: Do Diets Work?



The Atkins Diet is one of many popular weight loss diets. It is based on reducing the consumption of carbohydrates. For years, such “low-carb” diets have been touted as being effective for weight loss and other health benefits. But before 2001, no one had attempted to demonstrate the effectiveness of a low-carb diet in a well-designed comparative experiment. Then, two separate groups of researchers attempted to do just that.

At Duke University Medical Center, Dr. William Yancy and his colleagues recruited 120 people between the ages of 18 and 65. All of the participants were obese and had high cholesterol, but were otherwise in generally good health. Researchers randomly assigned half of the participants to a low-carbohydrate, high-protein diet (similar to an Atkins Diet) and the other half to a low-fat, low-cholesterol diet. At the end of six months, researchers measured the change in each participant’s weight and cholesterol levels.<sup>1</sup>

In the second study, Dr. Linda Stern and her colleagues recruited 132 obese adults at the Philadelphia Veterans Affairs Medical Center in Pennsylvania. Half of the participants were randomly assigned to a low-carbohydrate diet and the other half were assigned to a low-fat diet. Researchers measured each participant’s change in weight and cholesterol level after six months and again after one year.<sup>2</sup>

1. Complete the following table using the details provided above about the two studies.

	Duke University Study	Philadelphia Study
Subjects		
Factor(s)/Explanatory Variable(s)		
Treatments		
Response Variable(s)		

1 “A Low-Carbohydrate, Ketogenic Diet versus a Low-Fat Diet To Treat Obesity and Hyperlipidemia,” by Yancy, William S. et al, *Annals of Internal Medicine*, May 2004, 140(10) 769-777.

2 “The Effects of Low-Carbohydrate versus Conventional Weight Loss Diets in Severely Obese Adults: One-Year Follow-up of a Randomized Trial,” by Stern, Linda et al, *Annals of Internal Medicine*, May 2004, 140(10) 778-785.

2. Explain why both of these studies are experiments, and not observational studies or surveys.

3. How did the researchers in both studies determine which subjects received which treatments? Why did they use the method they did?

4. Could these experiments have been carried out in a single-blind or double-blind manner? Justify your answer.

5. Each of the following quotations describes the subjects in the Duke University experiment. Explain how each is an example of control and why it is important in terms of the design of the study.

(a) "None had dieted or used weight loss medications in the previous six months."

(b) "All subjects were encouraged to exercise 30 minutes at least three times per week and had regular group meetings at an outpatient research clinic for six months."

Let's look at some results from the two studies.

In the Duke University experiment, over the six-month duration of the study, weight loss was 12.9% of original body weight in the low-carbohydrate diet group and 6.7% of original body weight in the low-fat diet group. The low-carb diet group showed a greater increase in HDL (good) cholesterol than the low-fat diet group.

In the Philadelphia experiment, subjects in the low-carbohydrate diet group lost significantly more weight than subjects in the low-fat diet group during the first six months of the study. At the end of a year, however, the average weight loss for subjects in the two groups was not significantly different. The low-carbohydrate diet group did show greater increase in HDL (good) cholesterol level after a year than the low-fat diet group.

**6.** Briefly summarize what the results of these two experiments seem to suggest about the relative effectiveness of low-carbohydrate diets and low-fat diets on weight and cholesterol.

**7.** In the Philadelphia experiment, the subjects in the low-carbohydrate diet group lost an average of 5.1 kg in a year. The subjects in the low-fat diet group lost an average of 3.1 kg. Explain how this information could be consistent with the statement above about the average weight loss in the two groups not being significantly different.

**8.** Here is an excerpt from a report about the Duke University experiment: "Participants in the low-carbohydrate diet group had more minor adverse effects, such as constipation and headaches, than did patients in the low-fat diet group." How would you modify your summary in question 6 based on this additional information?



When you look at experimental results, it's important to consider possible limitations of the study. The next few questions will help you look critically at the two experiments described earlier.

**9.** Explain how the following excerpts from a report about the two experiments might affect your conclusions about the effectiveness of low-carb versus low-fat diets:

**Duke University study:** “The study was completed by 76% of participants in the low-carbohydrate diet group and by 57% of participants in the low-fat diet group.”

**Philadelphia study:** “Study limitations include high dropout rate of 34% ...”

**10.** In both experiments, participants were assigned at random to a low-fat or low-carbohydrate diet group. What exactly does that mean? The subjects in the low-fat diet group attended counseling sessions about how to restrict their caloric intake from fat. The subjects in the low-carbohydrate group attended counseling sessions about how to restrict their carbohydrate intake. These counseling sessions continued on a weekly or monthly basis throughout the experiment. It is possible that some people in each group did not restrict their diets as instructed. How might this affect conclusions based on the experiment?

**11.** In the Duke University study, subjects in the low-carbohydrate group all received daily nutritional supplements. Subjects in the low-fat group did not. How might this affect conclusions based on the experiment?

**12.** Give an example of a potential confounding variable in one of the two experiments. Explain carefully how the factor you choose could result in confounding.

**13.** Is it reasonable to generalize the results of these two experiments to the population of all overweight adults? Justify your answer.

**14.** Now that you have considered possible limitations of these two experiments, summarize what the results of these two experiments seem to suggest about the relative effectiveness of low-carbohydrate diets and low-fat diets on weight and cholesterol. You may want to refer to what you wrote earlier in response to question 6.



## Investigation #11: Distracted Learning



While you study, do you watch TV, listen to music, check your MySpace page, surf the Internet, chat on e-mail, talk or text on your cell phone? Do your parents insist that you can't possibly concentrate on studying while you're distracted by one of these activities? Maybe the conversation goes something like this:

**Parent:** "Take off your headphones and do your homework!"

**Student:** "I am doing my homework, and I work better with my music on."

**Parent:** "Turn it off! You can't study with that distraction!"

**Student:** "Yes I can. It helps me relax."

**Parent:** "Turn off that racket and concentrate on your school work!"

**Student:** "I study better with it on!"

Who is right? Some say that any distraction might interfere with your focus on the work you're doing, which may in turn affect the quality of the finished product. But others argue that listening to music actually helps them concentrate because the music "drowns out" other potential distractions. What do you think? Can previous research help us sort this out?<sup>1</sup>

In 1993, Frances Raucher and his colleagues designed an experiment to test whether listening to Mozart would help students improve their performance on a spatial reasoning task. They recruited 36 college students to participate in the experiment. The subjects were randomly assigned to three groups, with 12 students per group. Subjects in Group 1 listened to a 10-minute selection from a Mozart piece. Group 2 listened to a relaxation tape for 10 minutes. Subjects in Group 3 sat in silence for 10 minutes. Each subject took a pretest on spatial reasoning two days before the experiment and a post-test on spatial reasoning immediately after the 10-minute treatment. The results of the experiment seemed surprising: Students who listened to Mozart showed significantly higher gains in their scores on spatial-reasoning tasks than students in the other two groups.

After hearing the results of Rauscher's experiment, some eager parents started playing Mozart tapes for their children in hopes of increasing their spatial reasoning skills. One state even passed legislation requiring preschools to play 30 minutes of classical music a day. Other researchers tried to confirm this so-called "Mozart effect" in experiments of their own, but with little success.

So the question remains: Does listening to music help or hinder students' learning? The answer may depend on what type of "learning" we mean. In this investigation, your class will design and carry out an experiment to test whether listening to music

---

1 [www.madsci.org/posts/archives/mar98/889467626.Ns.r.html](http://www.madsci.org/posts/archives/mar98/889467626.Ns.r.html) served as inspiration for part of this investigation.

helps or hinders students as they perform a memorization task. Then, you will analyze data from the experiment and draw some preliminary conclusions from your research.

**1.** For simplicity, the members of your class will serve as the subjects in your experiment. How might this affect your ability to generalize the results of your study?

**2.** One possible design for the experiment would be to randomly assign about half of the students in your class to perform the memorization task while listening to Mozart, and the other half to perform the task in a silent room nearby. Then, you could compare the scores of students who listened to Mozart while memorizing with the scores of students who didn't. What flaw(s) do you see in using this design to conduct the experiment?

**3.** Some people are better at memorizing things than others. Here's another possible design for your experiment that takes this fact into account. Begin by having each student perform a memory task. Based on students' performance on this task, split the class into two roughly equal-sized groups containing the "good memorizers" and the "not-so-good memorizers." Randomly assign about half of the good memorizers to perform a second memory task while listening to Mozart, and the other half to perform the task in a silent room nearby. Use the same random assignment strategy for the not-so-good memorizers. To analyze the data from the experiment, you would compare the change in scores from the first memory task to the second for the good memorizers who listened to Mozart and those who didn't, and separately for the not-so-good memorizers who did and didn't listen to Mozart while memorizing.

(a) In what ways does this design improve on the design from question 2?

(b) How might you further improve the design of this experiment using the idea that some people are better memorizers than others? Explain.

4. Perhaps the best way to take individual differences in memorization skills into account in this experiment is to have each person perform two memory tasks—one while listening to Mozart and one in silence. Then, you can analyze data on the difference in performance for all students in your class and determine whether listening to Mozart seems to help or hurt memorization.

To carry out the experiment in this way, you will need two different but similar memory tasks. Let's call them task A and task B.

(a) Explain why you should not have all students perform task A while listening to Mozart and task B while in a silent room.

(b) Explain why you should not have all students perform their first memory task while sitting in a silent room and their second memory task while listening to Mozart, or vice versa.

(c) Discuss with your classmates how you could use random assignment to most effectively address the issues raised in parts (a) and (b). Once you have settled on a plan, propose it to your teacher.

(d) Describe carefully how you will perform the random assignment required by your approved plan from part (c).

**5.** Now that we have settled on a design for the experiment, let's confirm some of the details.

(a) Who are the subjects in this experiment?

(b) What factor(s)/explanatory variable(s) is this experiment investigating?

(c) What treatments are being administered? Explain why task A and task B are not treatments.

(d) Let's take a look at the tasks. Each subject will be presented with a list of 20 randomly generated two-digit numbers, such as the list shown below. The student will then have one minute to memorize as many of the numbers in the list as possible. At the end of the minute, each student will have two minutes to write down as many of the numbers as he or she can remember.

26      86      64      65      75      11      49      47      85      19  
 23      57      97      00      62      43      66      94      79      50

A wily student might just write down a bunch of two-digit numbers during the two minute period, hoping to match as many as possible. How might you score performance on this task to reward students for actual memorization and not for guessing?

(e) Based on your answer to (d), describe the response variable(s) this experiment will measure.

Now it's time to do the experiment! Your teacher will assist with logistics so that all students can participate.

**6.** Carry out the random assignment required for your experiment from question 4(d). Indicate clearly what each student will be doing first and second. You may find it helpful to make a chart like the one below that summarizes how the experiment will be carried out.

Subject	First Task	First Treatment	Second Task	Second Treatment
1	A	Music	B	Silence
2	A	Silence	B	Music
3	B	Music	A	Silence
4	B	Silence	A	Music





**7.** Have students perform the two memorization tasks as specified in question 6. Record data from the experiment in the table on the previous page.

**8.** Construct comparative dotplots or boxplots of the scores with music and the scores without music. Describe any similarities and differences you see in a few sentences.

**9.** Calculate the difference in scores for each student when listening to Mozart versus sitting in a silent room. As a class, decide on which order you will subtract the values. Record these values in the right-most column of the table on the previous page.

**10.** Construct an appropriate graph of the difference in memorization scores. Describe what the graph tells you in a couple of sentences.

**11.** In what way is the graph you constructed for question 10 more informative than the comparative graph from question 8?

**12.** Calculate a measure of center (mean or median) and a measure of spread that you think summarize the differences well. Explain why you chose the measures you did.

**13.** Was this experiment single-blind, double-blind, or neither? Justify your answer.

**14.** Based on the results of your experiment, does it appear that listening to Mozart helps or hinders students' performance on memorization tasks? Give appropriate graphical and numerical evidence to support your answer.

**15.** Can we generalize the results of this experiment to any kind of task that requires memorization? Justify your answer.

**16.** Why did we have all students listen to the same piece of Mozart music, rather than letting each student choose music he or she liked? Explain.



## Investigation #12: Would You Drink Blue Soda?



Does what you see affect your perception of how it tastes? If color can influence how people think a food tastes, what implications does this have for companies that make and market food and beverages?<sup>1</sup>

PepsiCo might be interested in your answer to these questions, as they have had two marketing failures based on introducing nontraditional colored beverages. In the early 1990s, PepsiCo introduced Pepsi Clear, a cola-flavored drink that was clear instead of brown in color. Pepsi Clear was later discontinued because sales were low. In 2002, PepsiCo tried again with Pepsi Blue.<sup>2</sup> Pepsi Blue was a berry-flavored cola drink that was blue in color. The Pepsi web site ([www.pepsi.com](http://www.pepsi.com)) says that Pepsi Blue was “created by and for teens. Through nine months of research and development, Pepsi asked young consumers what they want most in a new cola. Their response: Make it berry and make it blue.”

Unfortunately for PepsiCo, Pepsi Blue, like Pepsi Clear, was not a successful product, and it was discontinued a few years later. So what happened? Was the mistake adding a berry flavoring to cola, making the cola blue, or a combination of both?

In this investigation, you’ll investigate whether teens have a preference for or a dislike for blue-colored soda.

### Getting Started

To decide whether coloring a soda blue is a good or bad strategy if the drink is going to be marketed to teenagers, you will design and conduct an experiment, collect and analyze the data, and then make a recommendation.

For this experiment, you can start with a clear-colored soda, such as 7-Up or Sprite. Experiment with adding blue food coloring to the soda to create a “recipe” for a blue version of the soda. Food coloring is tasteless, so the addition of food coloring will not change the actual taste of the soda.

Once you have developed your new product, think carefully about how you would design an experiment to determine if teens have a preference for the clear soda or the blue soda.

**Note:** Be sure to discuss the ethical considerations involved in performing an experiment with human subjects. Your teacher will require you to obtain informed consent from all students (and possibly their parents) before they can participate in your experiment.

Once you have a plan in mind, answer the following questions. Be as specific as possible in your answers. It is OK to modify the design of your experiment if any of these

---

1 The page titled “Does the Color of Foods and Drinks Affect the Sense of Taste?” on the Neuroscience for Kids web site, <http://faculty.washington.edu/chudler/coltaste.html>, has a list of references to studies that have examined how color affects perceived taste.

2 You can find an announcement describing the launch of Pepsi Blue at <http://money.cnn.com/2002/05/07/news/companies/pepsi>.

questions reveal a weakness in your original plan. Now is the time to revise, before you actually carry out the experiment and collect the data!

**1.** In taste test experiments like the one you are designing, it is usual to randomize the order in which subjects taste the two drinks. That is, some subjects should taste the clear drink first and then the blue drink, while others should taste the blue drink first and then the clear. A random mechanism would be used to determine the order for each subject. Why do you think it is important to randomize the order in which the drinks are presented in an experiment of this type?

**2.** What would be a good way to determine the order (clear then blue or blue then clear) for each subject?

**3.** What are the two treatments for this experiment? *Hint:* In an experiment, subjects are assigned at random to one of the treatments.

**4.** Explain why it is not possible in this experiment to “blind” the subjects with respect to which experimental group they are in.

**5.** How will you select the subjects for your experiment, and how many subjects will participate? Be specific!

**6.** To what group, if any, will you be able to generalize the results of your experiment? Explain why you think it is reasonable to generalize to this particular group.

**7.** What question will you ask each subject after he or she has tasted the two sodas? Make sure that you will be able to determine from the response which of the two drinks was preferred.

**8.** After considering your answers to questions 1 through 7 and modifying your plan as needed, write a summary of your plan for conducting the experiment on *separate paper*. Include enough detail that someone who has not been part of your design team could read the summary and be able to carry out the experiment as you intended. Be sure to address ethical issues of using human subjects.

After your teacher has approved your experimental plan, carry out the experiment and collect data. Be sure to record the order in which the two drinks were tasted and the response for each subject.

Once you have collected the data, use it to fill in the four cells of the table below.

		Order	
		Clear then Blue	Blue then Clear
Preference	Clear		
	Blue		

9. Construct a graphical display that allows you to compare the preferences for the two experimental groups (clear then blue and blue then clear).

10. Based on your display, do you think there is a difference in preference for the two experimental groups? That is, do you think the order in which the drinks were tasted makes a difference? Explain.

11. Based on the data from this experiment, do you think there is a preference for one of the drinks (clear or blue) over the other? Explain, justifying your answer using the data from the experiment.

**12.** Write a report that makes recommendations to a soft drink company that is considering introducing a blue soft drink that will be marketed to teens. Include appropriate data and graphs to support your recommendations.