This article describes many of the different ways of designing a research project, as well as the different kinds of research that can be conducted. Please keep in mind as you read this that it was written by and for social researchers, who normally are very interested in studying people and the factors (variables) that affect people. As an engineer, you are equally likely, perhaps even more likely, to be interested in studying devices, materials, hardware, new techniques, and so on... These kinds of research can be classified and understood equally well using the concepts in this paper.

~ Dr. Seth P. Bates

Types of Research Designs

What are the different major types of research designs? We can classify designs into a simple threefold classification by asking some key questions. First, does the design use random assignment to groups? [Don't forget that random assignment is not the same thing as random selection of a sample from a population!] If random assignment is used, we call the design a randomized experiment or true experiment. If random assignment is not used, then we have to ask a second question: Does the design use either multiple groups or multiple waves of measurement? If the answer is yes, we would label it a quasi-experimental design. If no, we would call it a non-experimental design.

This threefold classification is especially useful for describing the design with respect to internal validity. A randomized experiment generally is the strongest of the three designs when your interest is in establishing a cause-effect relationship. A non-experiment is generally the weakest in this respect. I have to hasten to add here, that I don't mean that a non-experiment is the weakest of the three designs overall, but only with respect to internal validity or causal assessment. In fact, the simplest form of non-experiment is a one-shot survey design that consists of nothing but a single observation O. This is probably one of the most common forms of research and, for some research questions -- especially descriptive ones -- is clearly a strong design. When I say that the non-experiment is the weakest with respect to internal validity, all I mean is that it isn't a particularly good method for assessing the cause-effect relationship that you think might exist between a program and its outcomes.
To illustrate the different types of designs, consider one of each in design notation. The first design is a posttest-only randomized experiment. You can tell it's a randomized experiment because it has an R at the beginning of each line, indicating random assignment. The second design is a pre-post nonequivalent groups quasi-experiment. We know it's not a randomized experiment because random assignment wasn't used. And we know it's not a non-experiment because there are both multiple groups and multiple waves of measurement. That means it must be a quasi-experiment. We add the label "nonequivalent" because in this design we do not explicitly control the assignment and the groups may be nonequivalent or not similar to each other (see nonequivalent group designs).

Finally, we show a posttest-only nonexperimental design. You might use this design if you want to study the effects of a natural disaster like a flood or tornado and you want to do so by interviewing survivors. Notice that in this design, you don't have a comparison group (e.g., interview in a town down the road that didn't have the tornado to see what differences the tornado caused) and you don't have multiple waves of measurement (e.g., a pre-tornado level of how people in the ravaged town were doing before the disaster). Does it make sense to do the non-experimental study? Of course! You could gain lots of valuable information by well-conducted post-disaster interviews. But you may have a hard time establishing which of the things you observed are due to the disaster rather than to other factors like the peculiarities of the town or pre-disaster characteristics.
Experimental Design
Experimental designs are often touted as the most "rigorous" of all research designs or, as the "gold standard" against which all other designs are judged. In one sense, they probably are. If you can implement an experimental design well (and that is a big "if" indeed), then the experiment is probably the strongest design with respect to internal validity. Why? Recall that internal validity is at the center of all causal or cause-effect inferences. When you want to determine whether some program or treatment causes some outcome or outcomes to occur, then you are interested in having strong internal validity. Essentially, you want to assess the proposition:

If X, then Y

or, in more colloquial terms:

If the program is given, then the outcome occurs

Unfortunately, it's not enough just to show that when the program or treatment occurs the expected outcome also happens. That's because there may be lots of reasons, other than the program, for why you observed the outcome. To really show that there is a causal relationship, you have to simultaneously address the two propositions:

If X, then Y

and

If not X, then not Y

Or, once again more colloquially:

If the program is given, then the outcome occurs

and

If the program is not given, then the outcome does not occur

If you are able to provide evidence for both of these propositions, then you've in effect isolated the program from all of the other potential causes of the outcome. You've shown that when the program is present the outcome occurs and when it's not present, the outcome doesn't occur. That points to the causal effectiveness of the program.

Think of all this like a fork in the road. Down one path, you implement the program and observe the outcome. Down the other path, you don't implement the program and the outcome doesn't occur. But, how do we take both paths in the road in the same study? How can we be in two places at once? Ideally, what we want is to have the same conditions -- the same people, context, time, and so on -- and see whether when the program is given we get the outcome and when the
program is not given we don't. Obviously, we can never achieve this hypothetical situation. If we give the program to a group of people, we can't simultaneously not give it! So, how do we get out of this apparent dilemma?

Perhaps we just need to think about the problem a little differently. What if we could create two groups or contexts that are as similar as we can possibly make them? If we could be confident that the two situations are comparable, then we could administer our program in one (and see if the outcome occurs) and not give the program in the other (and see if the outcome doesn't occur). And, if the two contexts are comparable, then this is like taking both forks in the road simultaneously! We can have our cake and eat it too, so to speak.

That's exactly what an experimental design tries to achieve. In the simplest type of experiment, we create two groups that are "equivalent" to each other. One group (the program or treatment group) gets the program and the other group (the comparison or control group) does not. In all other respects, the groups are treated the same. They have similar people, live in similar contexts, have similar backgrounds, and so on. Now, if we observe differences in outcomes between these two groups, then the differences must be due to the only thing that differs between them -- that one got the program and the other didn't.

OK, so how do we create two groups that are "equivalent"? The approach used in experimental design is to assign people randomly from a common pool of people into the two groups. The experiment relies on this idea of random assignment to groups as the basis for obtaining two groups that are similar. Then, we give one the program or treatment and we don't give it to the other. We observe the same outcomes in both groups.

The key to the success of the experiment is in the random assignment. In fact, even with random assignment we never expect that the groups we create will be exactly the same. How could they be, when they are made up of different people? We rely on the idea of probability and assume that the two groups are "probabilistically equivalent" or equivalent within known probabilistic ranges. So, if we randomly assign people to two groups, and we have enough people in our study to achieve the desired probabilistic equivalence, then we may consider the experiment to be strong in internal validity and we probably have a good shot at assessing whether the program causes the outcome(s).

But there are lots of things that can go wrong. We may not have a large enough sample. Or, we may have people who refuse to participate in our study or who drop out part way through. Or, we may be challenged successfully on ethical grounds (after all, in order to use this approach we have to deny the program to some people who might be equally deserving of it as others). Or, we may get resistance from the staff in our study who would like some of their "favorite" people to get the program. Or, they mayor might insist that her daughter be put into the new program in an educational study because it may mean she'll get better grades.

The bottom line here is that experimental design is intrusive and difficult to carry out in most real world contexts. And, because an experiment is often an intrusion, you are to some extent setting up an artificial situation so that you can assess your causal relationship with high internal validity. If so, then you are limiting the degree to which you can generalize your results to real contexts where you haven't set up an experiment. That is, you have reduced your external validity in order to achieve greater internal validity.
In the end, there is just no simple answer (no matter what anyone tells you!). If the situation is right, an experiment can be a very strong design to use. But it isn't automatically so. My own personal guess is that randomized experiments are probably appropriate in no more than 10% of the social research studies that attempt to assess causal relationships.

Experimental design is a fairly complex subject in its own right. I've been discussing the simplest of experimental designs -- a two-group program versus comparison group design. But there are lots of experimental design variations that attempt to accomplish different things or solve different problems. In this section you'll explore the basic design and then learn some of the principles behind the major variations.

**Two-Group Experimental Designs**

The simplest of all experimental designs is the two-group posttest-only randomized experiment. In design notation, it has two lines -- one for each group -- with an R at the beginning of each line to indicate that the groups were randomly assigned. One group gets the treatment or program (the X) and the other group is the comparison group and doesn't get the program (note that this you could alternatively have the comparison group receive the standard or typical treatment, in which case this study would be a relative comparison).

Notice that a pretest is not required for this design. Usually we include a pretest in order to determine whether groups are comparable prior to the program, but because we are using random assignment we can assume that the two groups are **probabilistically equivalent** to begin with and the pretest is not required (although you'll see with covariance designs that a pretest may still be desirable in this context).

In this design, we are most interested in determining whether the two groups are different after the program. Typically we measure the groups on one or more measures (the Os in notation) and we compare them by testing for the differences between the means using a **t-test** or one way **Analysis of Variance (ANOVA)**.

The posttest-only randomized experiment is strong against the **single-group threats** to internal validity because it's not a single group design! (Tricky, huh?) It's strong against all of the **multiple-group threats** except for selection-mortality. For instance, it's strong against selection-testing and selection-instrumentation because it doesn't use repeated measurement. The selection-mortality threat is especially salient if there are differential rates of dropouts in the two groups. This could result if the treatment or program is a noxious or negative one (e.g., a painful medical procedure like chemotherapy) or if the control group condition is
painful or intolerable. This design is susceptible to all of the social interaction threats to internal validity. Because the design requires random assignment, in some institutional settings (e.g., schools) it is more likely to utilize persons who would be aware of each other and of the conditions they've been assigned to.

The posttest-only randomized experimental design is, despite its simple structure, one of the best research designs for assessing cause-effect relationships. It is easy to execute and, because it uses only a posttest, is relatively inexpensive. But there are many variations on this simple experimental design. You can begin to explore these by looking at how we classify the various experimental designs.

Probabilistic Equivalence

What is Probabilistic Equivalence?

What do I mean by the term probabilistic equivalence? Well, to begin with, I certainly don't mean that two groups are equal to each other. When we deal with human beings it is impossible to ever say that any two individuals or groups are equal or equivalent. Clearly the important term in the phrase is "probabilistic". This means that the type of equivalence we have is based on the notion of probabilities. In more concrete terms, probabilistic equivalence means that we know perfectly the odds that we will find a difference between two groups. Notice, it doesn't mean that the means of the two groups will be equal. It just means that we know the odds that they won't be equal. The figure shows two groups, one having a mean of 49 and the other with a mean of 51. Could these two groups be probabilistically equivalent? Certainly!

We achieve probabilistic equivalence through the mechanism of random assignment to groups. When we randomly assign to groups, we can calculate the chance that the two groups will differ just because of the random assignment (i.e., by chance alone). Let's say we are assigning a group of first grade students to two groups. Further, let's assume that the average test scores for these children for a standardized test with a population mean of 50 were 49 and 51 respectively. We might conduct a t-test to see if the means of our two randomly assigned groups are statistically

With $\alpha = .05$, we expect that we will observe a pretest difference 5 times out of 100

We achieve probabilistic equivalence through the mechanism of random assignment to groups. When we randomly assign to groups, we can calculate the chance that the two groups will differ just because of the random assignment (i.e., by chance alone). Let's say we are assigning a group of first grade students to two groups. Further, let's assume that the average test scores for these children for a standardized test with a population mean of 50 were 49 and 51 respectively. We might conduct a t-test to see if the means of our two randomly assigned groups are statistically
different. We know -- through random assignment and the law of large numbers -- that the chance that they will be different is 5 out of 100 when we set our significance level to .05 (i.e., *alpha* = .05). In other words, 5 times out of every 100, when we randomly assign two groups, we can expect to get a significant difference at the .05 level of significance.

When we assign randomly, the only reason the groups can differ is because of chance assignment because their assignment is entirely based on the randomness of assignment. If, by chance, the groups differ on one variable, we have no reason to believe that they will automatically be different on any other. Even if we find that the groups differ on a pretest, we have no reason to suspect that they will differ on a posttest. Why? Because their pretest difference had to be a chance one. So, when we randomly assign, we are able to assume that the groups do have a form of equivalence. We don't expect them to be equal. But we do expect that they are "probabilistically" equal.

**Random Selection & Assignment**

**Random selection** is how you draw the sample of people for your study from a population.

**Random assignment** is how you assign the sample that you draw to different groups or treatments in your study.

It is possible to have *both* random selection and assignment in a study. Let's say you drew a random sample of 100 clients from a population list of 1000 current clients of your organization. That is random sampling. Now, let's say you randomly assign 50 of these clients to get some new additional treatment and the other 50 to be controls. That's random assignment. It is also possible to have *only one of these* (random selection or random assignment) but not the other in a study. For instance, if you do not randomly draw the 100 cases from your list of 1000 but instead just take the first 100 on the list, you do not have random selection. But you could still randomly assign this nonrandom sample to treatment versus control. Or, you could randomly select 100 from your list of 1000 and then nonrandomly (haphazardly) assign them to treatment or control.

And, it's possible to have *neither* random selection nor random assignment. In a typical nonequivalent groups design in education you might nonrandomly choose two 5th grade classes to be in your study. This is nonrandom selection. Then, you could arbitrarily assign one to get the new educational program and the other to be the control. This is nonrandom (or nonequivalent) assignment.

Random selection is related to sampling. Therefore it is most related to the external validity (or generalizability) of your results. After all, we would randomly sample so that our research participants better represent the larger group from which they're drawn. Random assignment is most related to design. In fact, when we randomly assign participants to treatments we have, by definition, an experimental design. Therefore, random assignment is most related to internal validity. After all, we randomly assign in order to help assure that our treatment groups are similar to each other (i.e., equivalent) prior to the treatment.

---