**Some notes from Oehlert’s book**

**Experiments answer questions:**

• Is a drug a safe, effective cure for a disease? This could be a test of

how AZT affects the progress of AIDS.

• Which combination of protein and carbohydrate sources provides the

best nutrition for growing lambs?

• How will cellphone usage change if our company offers

a different rate structure to our customers?

• Will an ice cream manufactured with a new kind of stabilizer be as

palatable as our current ice cream?

• Does short-term incarceration of spouse abusers deter future assaults? (In Alabama)

• Under what conditions should I operate my chemical refinery, given

this month’s grade of raw material?

Chemical experiment: the reactivity of an [alkali metal](https://www.thoughtco.com/definition-of-alkali-metal-604361) (Li, Na, K, Rb, Cs, Fr) with water. A small piece of sodium metal will be placed in a bowl of water.

The reaction is:

2 Na + 2 H2O → 2 Na+ + 2 OH- + H2(g)

The reaction is especially vigorous when warm water is used

<https://www.youtube.com/watch?v=nWu44AqF0iI> Is this an experiment?

<https://www.youtube.com/watch?v=Bm9xXyw2f7g> Is this an experiment?

An experiment is characterized by the treatments and experimental units to

be used, the way treatments are assigned to units, and the responses that are

measured.

Experiments help us answer questions, but there are also nonexperimental techniques. What is so special about experiments? Consider that:

1. Experiments allow us to set up a direct comparison between the treatments

of interest.

2. We can design experiments to minimize any bias in the comparison.

3. We can design experiments so that the error in the comparison is small.

4. Most important, we are in control of experiments, and having that control

allows us to make stronger inferences about the nature of differences

that we see in the experiment. Specifically, we may make inferences

about *causation*.

This last point distinguishes an experiment from an *observational study*

An observation observational study also has treatments, units, and responses. However, in the observational study we merely observe which units are in which treatment groups; we don’t get to control that assignment.

Observational studies are useful too. It is important to say that while experiments have some advantages, observational studies are also useful and can produce important results. For example, studies of smoking and human health are observational, but the link that they have established is one of the most important public health issues today. Same idea goes with vaping studies.

Mosteller and Tukey (1977) list three concepts associated with causation and state that two or three are needed to support a causal relationship:

*• Consistency*

*• Responsiveness*

*• Mechanism*.

Consistency means that, all other things being equal, the relationship between two variables is consistent across populations in direction and maybe in amount. Responsiveness means that we can go into a system, change the causal variable, and watch the response variable change accordingly. Mechanism means that we have a step-by-step mechanism leading from cause to effect.

Experiments can demonstrate Consistency and Responsiveness.

A good experimental design must

*• Avoid systematic error*

*• Be precise*

*• Allow estimation of error*

*• Have broad validity.*

For example, if responses for units receiving treatment systematic error one are measured with instrument A, and responses for treatment two are measured with instrument B, then we don’t know if any observed differences are due to treatment effects or instrument miscalibrations (this is called *confounding* and can be diminished by using *randomization*).

Precision means smaller random errors. High precision means low variability in the response.

The error cannot be estimated with just one observation of the treatment effect.

The conclusions we draw from an experiment are applicable to the experimental units we used in the experiment. If the units are actually a statistical sample from some population of units, then the conclusions are also valid for the population. Beyond this, we are extrapolating, and the extrapolation validity might or might not be successful. For example, suppose we compare two different drugs for treating attention deficit disorder. Our subjects are preadolescent boys from our clinic. We might have a fair case that our results would hold for preadolescent boys elsewhere, but even that might not be true if our clinic’s population of subjects is unusual in some way. The results are even less compelling for older boys or for girls. Thus, if we wish to have wide validity—for example, broad age range and both genders—then our experimental units should reflect the population about which we wish to draw inference.

In practice we may need to compromise.

TERMS AND CONCEPTS IN DOX

**Treatments** are the different procedures we want to compare. These could be different kinds or amounts of fertilizer in agronomy, different long distance rate structures in marketing, or different temperatures in a reactor vessel in chemical engineering.

**Experimental Units** are the things to which we apply the treatments. These could be plots of land receiving fertilizer, groups of customers receiving different rate structures, or batches of feedstock processing at different temperatures.

**Responses** are outcomes that we observe after applying a treatment to an experimental unit. That is, the response is what we measure to judge what happened in the experiment; we often have more than one response. Responses might be nitrogen content or biomass of corn plants, profit by customer group, or yield and quality of the product per ton of raw material. In general the GOAL of an experiment is to have a *low response* or a *high response* or a *response close to a target value.*

**Randomization** is the use of a known, understood probabilistic mechanism for the assignment of treatments to units. Other aspects of an experiment can also be randomized: for example, the order in which units are evaluated for their responses.

**Experimental Error** is the random variation present in all experimental results. Different experimental units will give different responses to the same treatment, and it is often true that applying the same treatment over and over again to the same unit will result in different responses in different trials. Experimental error does not refer to conducting the wrong experiment or dropping test tubes.

**Measurement units (or response units)** are the actual objects on which the response is measured. These may differ from the experimental units. For example, consider the effect of different fertilizers on the nitrogen content of corn plants. Different field plots are the experimental units, but the measurement units might be a subset of the corn plants on the field plot, or a sample of leaves, stalks, and roots from the field plot.

**Blinding** occurs when the evaluators of a response do not know which treatment was given to which unit. Blinding helps prevent bias in the evaluation, even unconscious bias from well-intentioned evaluators. Double blinding occurs when both the evaluators of the response and the (human subject) experimental units do not know the assignment of treatments to units. Blinding the subjects can also prevent bias, because subject responses can change when subjects have expectations for certain treatments.

**Control** has several different uses in design. First, an experiment is *controlled* because we as experimenters assign treatments to experimental units. Otherwise, we would have an observational study. Second, a control treatment is a “standard” treatment that is used as a *baseline* or basis of comparison for the other treatments. This control treatment might be the treatment in common use, or it might be a null treatment (no treatment at all). For example, a study of new pain killing drugs could use a standard pain killer as a control treatment, or a study on the efficacy of fertilizer could give some fields no fertilizer at all. This would control for average soil fertility or weather conditions.

**Placebo** is a null treatment that is used when the act of applying a treatment—any treatment—has an effect. Placebos are often used with human subjects, because people often respond to any treatment: for example, reduction in headache pain when given a sugar pill. Blinding is important when placebos are used with human subjects. Placebos are also useful for nonhuman subjects. The apparatus for spraying a field with a pesticide may compact the soil. Thus, we drive the pparatus over the field, without actually spraying, as a placebo treatment.

**Factors** combine to form treatments. For example, the baking treatment for a cake involves a given time at a given temperature. The treatment is the combination of time and temperature, but we can vary the time and temperature separately. Thus, we speak of a time factor and a temperature factor. Individual settings for each factor are called levels of the factor; other name is *treatment combination.*

**Confounding** occurs when the effect of one factor or treatment cannot be distinguished from that of another factor or treatment. The two factors or treatments are said to be confounded. Except in very special circumstances, confounding should be avoided. Consider planting tomato variety A in New Jersey and tomato variety B in Iowa. In this experiment, we cannot distinguish location effects from variety effects—the variety factor and the location factor are confounded.