Overview

This reading will discuss the differences between **between-subjects** and **within-subjects** independent variables and will discuss some issues that are specific to studies that use each type. Recall that another label for independent variable is "factor." A study that uses only between-subjects factors is said to use a **between-subjects design**, and a study that uses only within-subjects factors is called a **within-subjects design**. First, we will review the features of between-subjects factors and then we will contrast them with the features of within-subjects factors. Then we will discuss a way to combine the two types of factors to create a "**mixed**" design (one that has a mix of both between-subjects and within-subjects factors).

Between-Subjects Factors

The hallmark of a **between-subjects factor** is that each participant is assigned to one and only one level of each factor. For example, participants might be randomly assigned to either receive negative feedback or positive feedback. Feedback is the independent variable, and it has two levels: positive or negative. It is a between-subjects factor because each participant only receives one type of feedback. There are two independent groups of participants: one receiving positive feedback and the other receiving negative feedback.

In a between-subjects design, the typical approach to statistical analysis is to compare the means of the different levels of the between-subjects factor. To use the above example, we might measure each participant's self-esteem after he or she has received feedback. The mean self-esteem score for the positive feedback group would then be compared to the mean self-esteem score for the negative feedback group. Imagine you obtain the following scores from 10 participants:

Positive Feedback	*Negative Feedback*
46	35
41	38
39	41
44	37
48	42
mean = 43.6	mean = 38.6

Those 10 scores are not all the same; they *vary*. A researcher's goal is to explain as much of that variance as possible. The variance that you *can* explain is the variance due to being in the positive versus negative feedback condition. This is the variance between the means of the two groups: 43.6 versus 38.6. You can explain that variance because you have an independent variable – "feedback condition" – that distinguishes between those groups. However, there is also variance that you cannot explain. You cannot explain why one subject in the positive feedback condition has a self-esteem of 46 and another has a self-esteem of 41. You can't explain that difference because you do not have any variables that distinguish those two subjects; you don't have any information about how they differ from one another. That is "error variance": differences between scores that you cannot explain. In a between-subjects design, error variance is the variability in the scores *within each condition*. As the scores within each condition become more spread out, error variance increases. As error variance increases, it becomes harder to detect whether an effect exists (e.g., whether self-esteem is influenced by type of feedback). An important consequence of this is that researchers do what they can to reduce within-condition variability. How would they do that? First, by treating every subject within a condition as similarly as possible. Second, by seeking a **homogeneous sample**: a sample of people who are very similar to one another. Thus, reducing error variance and getting a diverse

sample are incompatible goals: increasing the diversity of your sample will also increase your error variance.

Within-Subjects Factors

The hallmark of a within-subjects factor is that the same subjects are exposed to more than one level of the factor. For example, if in our earlier example we had the same subjects receive both positive *and* negative feedback and we measured their self-esteem after each one, feedback condition would be a within-subjects factor:

Subject	*Positive Feedback*	*Negative Feedback*	*Difference*
1	46	42	4
2	41	37	4
3	39	36	3
4	44	39	5
5	48	44	4
variance[1]	13.3	11.3	0.5

The primary advantage of within-subjects factors over between-subjects factors is that within-subjects factors have greater **statistical power** than between-subjects factors. This means that within-subjects factors are better able to detect an effect, given that one exists. You can think of within-subjects factors as being like microscopes that have greater magnification: They allow you to detect tiny effects that would have gone unnoticed in between-subjects factors. The reason they have greater power is that they have smaller error variance than between-subjects factors. For a within-subjects factor, error variance is computed from the variance in the *difference* scores. Look in the table above under the "Difference" column. Note how for each subject, self-esteem is about 4 units higher when they receive positive feedback than when they receive negative feedback. However, there is variance in these differences. Some subjects show a slightly larger effect (subject 4) and some subjects show a slightly smaller effect (subject 3). This variance in the difference from subject to subject is variance you cannot explain: error variance. The variance[1] of the 5 difference scores is only 0.5. Compare that to the variance of the scores in the "Positive Feedback" or "Negative Feedback" conditions. Those are the error variances you would have from a between-subjects design. They are more than 20 times higher. In general, the error variance in a within-subjects design is much smaller than the error variance in a between-subjects design.

A common way to think about why within-subjects factors have more power than between-subjects factors is to think of *each subject serving as his or her own control group*. Instead of putting one person into the experimental condition and another person into a control condition, you are putting the same person into both conditions. Because it is the same person, the difference between their scores in the two conditions cannot be due to age, personality, or any other individual difference. You have removed all of those sources of error variance.

Another way to think about within-subjects factors' advantages is to consider how they *control for baseline differences among subjects in the dependent variable*. "Baseline" refers to the average level of a variable. In the example above, a person's baseline self-esteem would be their typical self-esteem level. This variable varies considerably from person to person. Some people will come into

1 Variance is computed by finding the mean of the scores, subtracting the mean from each score, squaring each of those, and adding them up. It is a way of measuring how spread out a group of numbers is. As scores get more spread out, variance increases, and as scores get more tightly packed together, variance decreases.

the study with low self-esteem and some will come into the study with high self-esteem. If feedback condition were a between-subjects factor, all of those pre-existing differences in self-esteem would contribute to error variance. However, if feedback condition were a *within*-subjects factor, the variance in pre-existing levels of self-esteem would be irrelevant because all you care about is how much each person *changes* from one condition to another. You don't care that one subject came in with a self-esteem of 46 and another came in with a self-esteem of 39. All you care about is how much their self-esteem scores *change* from the positive to the negative feedback conditions. Any time you are studying a dependent variable that shows lots of variability between people, you should consider using a within-subjects factor to minimize the error variance.

Although within-subjects factors have more statistical power, they also have a major limitation. By exposing subjects to all of the levels of your independent variable (e.g., both positive and negative feedback), you increase the risk that they will discover your hypothesis and respond differently because of it. This is more of a problem when subjects are (a) motivated and (b) able to change their behavior. For example, in a study on the effects of number of syllables on word memorization, subjects would have little motivation to modify their responses. In contrast, if subjects figured out that they were in a study on the effects of interviewer race on nonverbal indicators of anxiety, they might try to act more relaxed than they otherwise would so that they do not appear to be racist. In addition, there are several potential threats to internal validity for within-subjects factors. These threats are covered in a separate reading titled “Threats to Internal Validity for Within-Subjects Designs”.

Mixed Designs

When a study has at least one between-subjects factor *and* at least one within-subjects factor, it is said to have a “mixed” design. Let's begin with a common within-subjects factor: time. In a pre-post design, subjects are measured both before and after some treatment is applied. For example, a group of subjects with depression may have their sadness measured before they receive some medication and then again three weeks after taking the medication. The independent variable is time (at two levels: before and after) and the dependent variable is sadness. As discussed in the “Threats to Internal Validity for Within-Subjects Designs” reading, there is a big problem with that study. Any changes from pre to post may be due to naturally-occurring internal processes instead of the medication. To remove that concern, we can add a between-subjects factor: treatment, at two levels: medication or no medication. This would be described as a “2 (time: pre versus post) x 2 (treatment: medication versus no medication) mixed design, with repeated measures on the first factor”. “Repeated measures” is another way of saying “within-subjects”, and identifies which factor is within-subjects. The following table provides a way of representing the results from such a study. The numbers in each of the 4 cells represent the average values for that condition.

DV: sadness		*Time*	
		Pre	*Post*
Treatment	*Medication*	47.3	28.6
	No medication	48.1	46.5

Note how the change in sadness from pre to post is greater for the medication condition than it is for the no-medication condition. This suggests that the medication is doing something beyond the body's normal response. Many of the threats to internal validity for within-subjects factors can be reduced with the addition of a between-subjects factor, turning the study into a mixed design.

Another threat to validity for within-subjects factors is **order effects**: the difference between two tasks may be due to the order in which they were completed rather than the nature of the tasks themselves. For example, if subjects complete a word-scramble task while listening to music and then complete another word-scramble task with no music, the difference in their performance could be due

to the music *or* to the fact that one task was done first and the other was done second. Perhaps subjects get better with practice, and the improved scores in the no-music condition are actually due to practice rather than the absence of music. To rule this out, a researcher could make *order* a between-subjects factor. Half the subjects would have music first, and half the subjects would have silence first. This would be a 2 (music: present versus absent) by 2 (order: music first vs silence first) mixed design with repeated measures on the first factor. It could be represented like this:

DV: word-scramble performance		*Music*	
		Present	*Absent*
Order	*Music first*	78.4	85.6
	Silence first	78.6	85.4

In the above table, we can see that order does not appear to be an important factor because the pattern of results is the same regardless of whether music or silence came first. In both conditions, subjects did better when there was no music. However, by adding order as a between-subjects factor, at least we can now confirm that it has no effect rather than simply hoping that it does not.

Mixed designs combine the advantages of within-subjects and between-subjects factors. The within-subjects factors in a mixed design add statistical power, while the between-subjects factors help to rule out threats to internal validity. However, there is a special threat to internal validity for mixed designs. It is called **differential mortality** and it occurs when the dropout rate (called “mortality”, although it includes subjects just dropping out and not actually dying) is different across the levels of the between-subjects factor. For example, imagine a study on the effects of vigorous exercise on mood. Forty subjects are randomly assigned to either vigorous exercise or a control group who watches TV. Their mood is measured every week for 6 weeks. The table below shows the *number of subjects* across the six weeks.

Number of participants	*week*					
	1	*2*	*3*	*4*	*5*	*6*
Exercise	20	16	12	10	8	8
TV	20	20	20	19	19	18

Note how the dropout rate is much higher for one group (exercise) than the other (TV). That is differential mortality. When the subjects are randomly assigned to the two conditions in week 1, we can assume that those two groups are approximately equal on all important confounding variables. For example, they are equally energetic and extroverted. Differential mortality poses a risk because it destroys that assumption of equality. By week 6, you've only got 8 people left in the exercise group. How might the 8 people remaining in the exercise group be different from the 12 who dropped out? They probably enjoy exercise more. Any comparisons between the two conditions at week 6 would be compromised by selection effects: The groups probably differed before the study began. You would not be testing whether exercise makes people happier, you would be testing whether exercisers are happier than non-exercisers. Those exercisers may be happier not because they exercise but because of their personalities.

Summary

In a between-subjects factor, each subject is assigned to only one level. In a within-subjects factor, each subject is assigned to more than one level (and usually all levels). Error variance for between-subjects factors comes from the variance in scores within a condition, whereas error variance for within-subjects factors comes from the variance in difference scores between conditions. As a result, within-subjects factors tend to have lower error variance and therefore more statistical power. However, within-subjects factors are more vulnerable to subjects discovering the hypothesis and suffer from several other threats to internal validity. One way to reduce those vulnerabilities is to use a study that combines within-subjects and between-subjects factors, which is called a “mixed” design. Mixed designs combine the advantages of within-subjects factors (greater statistical power) and between-subjects factors (less risk of subjects discovering the hypothesis) but are vulnerable to differential mortality, a threat to internal validity caused by differing dropout rates among levels of the between-subjects variable.