Main Effects and Interactions

So far, we've talked about studies in which there is just one independent variable, such as "violence of television program." You might randomly assign people to watch television programs with either lots of violence or no violence and then compare them in some way, such as their attitudes toward the death penalty. We've also talked about studies that have more than just two levels of the independent variable. Using the example above, we could add a level in which people watched television programs with a moderate amount of violence. Even though there are three levels, there is still just one independent variable: TV violence. This chapter is designed to introduce you to studies where there is more than one independent variable. For example, you might be curious about whether the effect of TV violence is different for men and women. In this case, you would want to conduct a study with two independent variables: TV violence and gender.

Factorial Design

A study that has more than one independent variable is said to use a **factorial design**. A "factor" is another name for an independent variable. Factorial designs are described using "A x B" notation, in which "A" stands for the number of levels of one independent variable and "B" stands for the number of levels of the second independent variable. For example, if you are using two levels of TV violence (high vs. none) and two levels of gender (male vs. female), then you are using a 2 x 2 factorial design. If you add a medium level of TV violence to your design, then you have a 3 x 2 factorial design. In your methods section, you would write, "This study is a 3 (television violence: high, medium, or none) by 2 (gender: male or female) factorial design." A 2 x 2 x 2 factorial design is a design with three independent variables, each with two levels.

Main Effects

A "main effect" is the effect of one of your independent variables on the dependent variable, ignoring the effects of all other independent variables. To examine main effects, let's look at a study by Rosenthal and Jacobson (1966) in which first-graders and sixth-graders are given IQ tests, and then two weeks later, their teachers are given the names of 20% of the students in their class who were expected to show "unusual intellectual gains." These students were selected completely at random, without regard to their actual test scores, to see if teacher expectations alone have an impact on student performance. We include grade as another factor to see if teacher expectations have a different effect depending on the grade of the student. This would be a 2 (teacher expectations: high or normal) x 2 (grade: 1 or 6) factorial design. Eight months after the teachers are given high expectations for some students, all the students are given another IQ test. The mean IQ gains for the four possible conditions of this study, loosely based on Rosenthal and Jacobson (1966), are given below in Table 1.

Table 1

Mean Gain in IQ Scores by Teacher Expectation and Grade

Teacher expectations	Grade	
	First	Sixth
high	27.0	11.7
normal	11.6	10.2

Because a main effect is the effect of one independent variable on the dependent variable, ignoring the effects of other independent variables, you will have a total of two potential main effects in this study: one for grade of student and one for teacher expectations. **In general, there is one main effect for every independent variable in a study.** To look for a main effect of teacher expectations, you would calculate the average IQ score *across both first-graders and sixth-graders*. This is done in Table 2.

Table 2

Main Effect of Teacher Expectations

Teacher expectations	Grade		
	First	Sixth	***Average***
high	27.0	11.7	***19.33***
normal	11.6	10.2	***10.92***

Note that these averages assume that there are an equal number of people in the first-grade and the sixth-grade conditions[1]. Looking at these two averages, we see that they differ by 8.41 IQ points. Students whose teachers had high expectations scored, on average, 8.41 points higher than students whose teachers had normal expectations. To determine whether the “main effect of teacher expectation on IQ score” is significant, you would need to test whether the difference of 8.41 IQ points is greater than you would expect by chance. To do this, you need a statistical test. Before we get to that test, however, we should look at the main effect of grade.

Table 3

Main Effect of Age of Student

Teacher expectations	Grade	
	First	Sixth
high	27.0	11.7
normal	11.6	10.2
Average	***19.28***	***10.97***

In Table 3, we see that IQ scores of first-graders and sixth-graders differ by 8.31 points, on average, with first-graders gaining more than sixth-graders. To determine whether there is a main effect of grade, you would need to test whether the 8.31-point difference is greater than you would expect by chance.

Detecting main effects in Deducer. To analyze a factorial design in Deducer, you would select Analysis → ANOVA. You would then get the screen shown in Figure 1.

[1] Having unequal numbers across conditions can produce misleading results, so it is advisable to set the number of subjects per condition to be constant. In Rosenthal and Jacobson's (1966) original data, they had unequal numbers (43 in normal expectations, 11 in high, per grade). For simplicity, I have changed this so that each cell has 30 subjects.

Figure 1. Setting up the analysis of the effects of teacher expectations and grade on IQ score.

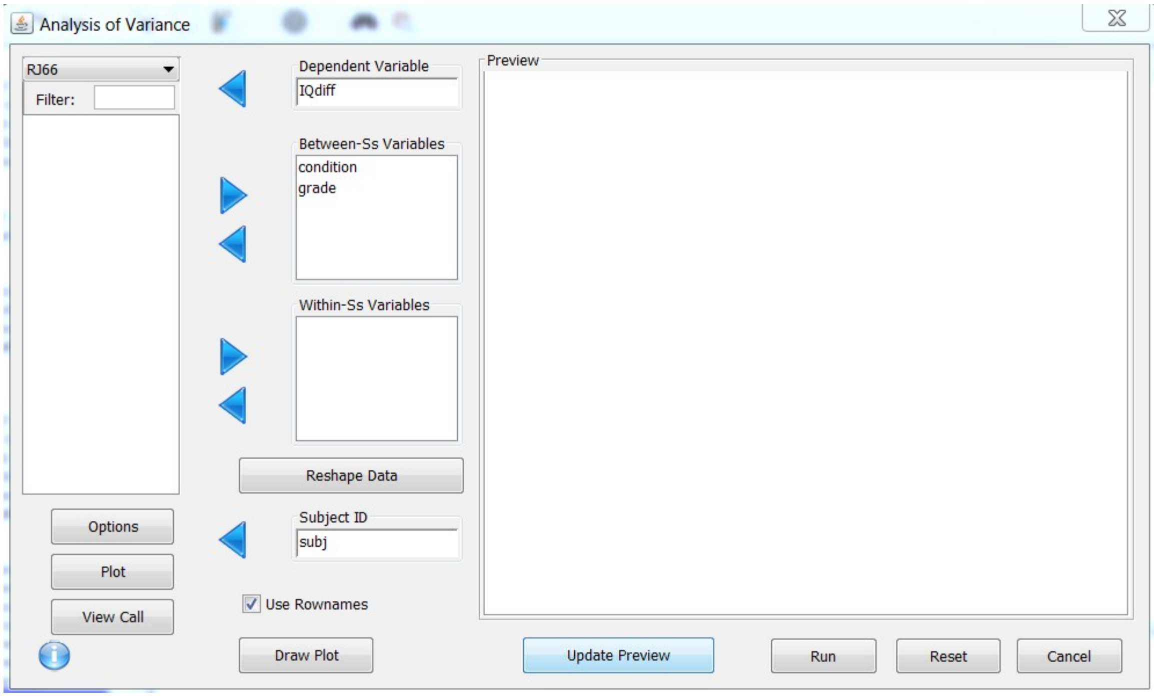


Figure 1 is similar to the layout we used for one-way ANOVA, but note that there are now two independent variables in the "Between-Subjects Variables" window. Clicking "Update Preview" produces the output shown in Figure 2.

Figure 2

Deducer Output from Analysis of Effect of Teacher Expectation and Grade on IQ Change

Preview: ANOVA | Levene's Test for Homogeneity of Variance | descriptives | Tukey

	Effect	DFn	DFd	F	p	p<.05	ges
2	condition	1	116	85.5	<0.001	*	0.424
3	grade	1	116	83.5	<0.001	*	0.419
4	condition:grade	1	116	58.3	<0.001	*	0.335

Significance and Effect Size

Figure 2 is an "ANOVA table." The test of the main effect of condition is shown in the first row. The *F* test statistic would be 1.0 if the two grades had exactly the same gains in IQ. The *p*-value tells you the probability of obtaining an *F* statistic that is 85.5 or higher just by chance: <.001. Because the usual alpha level in psychology is set to .05, you would conclude that the main effect of grade is statistically significant. An *F* of 85.5 or higher occurs by chance less than 1 time out of a thousand ($p < .001$). Statistical significance tells you that an effect is unlikely to be due to chance, but it does not tell you how large or meaningful the effect is. For that, you need a measure of effect size.

There are two ways to evaluate effect size: concrete units of the dependent variable, and standardized measures of effect size. The dependent variable in this example, change in IQ points, is in units that many people can understand, so it would be worthwhile to report effects in those units. As you may have learned from an introductory psychology course, IQ scores have a mean of 100 and a standard deviation of 15. As we computed above, the main effect of grade is 8.42 IQ points. Most people familiar with IQ scores would regard that as a meaningful difference but not an enormous difference.

Another way to report effect size is with a standardized measure, such as *d, r,* or *eta squared,* which were introduced in the reading on statistical inference. The last column in Figure 2 shows such a statistic: *ges,* for "generalized eta squared." This statistic does not have an easily interpretable meaning, but, as mentioned in the statistics exercise on one-way ANOVA, some benchmarks for *ges* are 0.02 for a small effect, 0.13 for a medium effect, and 0.26 for a large effect. Generalized eta squared is based on eta squared, which is the proportion of variance in the dependent variable that a factor can explain (see Bakeman, 2005, for more details). In the example above, all the effects would be classified as large in size because they are between 0.34 and 0.42.

Communicating Main Effects

If any main effect is significant, you must also report the pattern of means for that main effect. In Table 2, we saw that the main effect of teacher expectations was 8.41 points. When you report an estimate, it is useful to also report the confidence interval for that estimate so that your audience knows the precision of that estimate. To obtain the confidence interval for the main effects in Deducer, click on the "Tukey" tab in the Preview window.

Figure 3. Results of Tukey's HSD

```
Preview
ANOVA | Levene's Test for Homogeneity of Variance | descriptives | Tukey

$condition
               diff   lwr   upr  p adj
normal-high -8.42 -10.2 -6.61 <0.001

$grade
               diff   lwr   upr  p adj
sixth-first -8.32 -10.1 -6.51 <0.001

$`condition:grade`
                               diff    lwr    upr  p adj
normal:first-high:first   -15.37 -18.72 -12.01 <0.001
high:sixth-high:first     -15.27 -18.62 -11.91 <0.001
normal:sixth-high:first   -16.73 -20.09 -13.38 <0.001
high:sixth-normal:first     0.10  -3.25   3.45  1.000
normal:sixth-normal:first  -1.37  -4.72   1.99  0.713
normal:sixth-high:sixth    -1.47  -4.82   1.89  0.666
```

The values under the "diff" column should look familiar: they are the values we calculated in Tables 2 and 3 (with slight differences due to rounding). In addition, we have the lower and upper limits of the 95% confidence interval for those differences. Here's how it might look in APA style:

> The main effect of teacher expectation on IQ gain was significant, $F(1,116) = 85.5$, $p < .001$, *ges* = .42. Students whose teachers had high expectations gained 8.42 more IQ points than students whose teachers had normal expectations (95% *CI of the difference*

= 6.6 to 10.2 IQ points). The main effect of student grade on IQ gain was also significant, *F*(1,116) = 83.5, $p < .001$, *ges* = .42. First-graders gained 8.32 more IQ points than sixth-graders (95% *CI of the difference* = 6.5 to 10.1 IQ points).

The confidence intervals help to express the precision of your estimate. You can be 95% confident that the "true" effect of teacher expectations (defined as the effect you would obtain if you included the entire target population) is between 6.6 and 10.2 IQ points. There is a 5% risk that the true difference is outside that interval.

Two things to keep in mind when writing out results: 1) All the statistical letters (*F* and *p*) are italicized; and 2) The phrasing is of the format: "The main effect of the IV on the DV."

Interactions

A statistical interaction occurs when the effect of one independent variable on the dependent variable *changes* depending on the level of another independent variable. In Rosenthal and Jacobson's (1966) study, this is equivalent to asking whether the effect of teacher expectations changes depending on the grade of the student. If the effect of teacher expectations on IQ for first-graders is different from the effect of teacher expectations on IQ for sixth-graders, then there is an interaction. To determine if this is the case, we need to look at the **simple main effects** *(or just "simple effects")*: the main effect of one independent variable (e.g., teacher expectation) at each level of another independent variable (grade). This is shown in Table 4.

Table 4.

Simple Effects

Teacher expectations	Grade	
	First	Sixth
high	27.0	11.7
normal	11.6	10.2
Simple effect	***15.4***	***1.5***

The simple effect of teacher expectation for first-graders is 15.4 points, whereas the simple effect of teacher expectation for sixth-graders is 1.5 points. This suggests a statistical interaction: The effect of one IV (teacher expectation) is *changing* depending on another IV (grade). First-graders show a big effect of teacher expectations, but sixth-graders show almost no effect.

In the case of an interaction, the null hypothesis is that the simple effects are equal. For this example, a test of the significance of the interaction would be a test of whether the difference between 15.4 and 1.5 is so large that it is unlikely to have occurred by chance. This information is provided in the output in Table 4 in the row labeled "condition:grade". The condition by grade interaction is significant, $p < .001$, indicating that the difference in simple main effects is so large that it would be expected less than one time out of 1000 by chance. Now we know that the interaction is significant, but how do we interpret it?

To interpret a significant interaction, describe how the simple effects are different. The results from Tukey's HSD can be helpful here:

Figure 4. Tukey HSD Results for the Interaction (from bottom of Figure 3)

```
$`condition:grade`
                             diff    lwr    upr  p adj
normal:first-high:first    -15.37 -18.72 -12.01 <0.001
high:sixth-high:first      -15.27 -18.62 -11.91 <0.001
normal:sixth-high:first    -16.73 -20.09 -13.38 <0.001
high:sixth-normal:first      0.10  -3.25   3.45  1.000
normal:sixth-normal:first   -1.37  -4.72   1.99  0.713
normal:sixth-high:sixth     -1.47  -4.82   1.89  0.666
```

The Tukey HSD results provide more comparisons than you will usually be interested in. To find which row you should focus on, recall the interaction you are examining: how do the effects of expectation change across grade. That would lead you to focus on high-vs.-control for 1st graders, and then high-vs.-control for 6th graders: the first row and the last row.

Conceptually, the interaction can be explained like this:

> 1st-graders: high >> control (high is much higher than control)
>
> 6th-graders: high = control (high is approximately equal to control).

To write this in APA style, you could say:

> There was a significant interaction between grade and teacher expectations, *F*(1,116) = 58.3, *p* < .001, *ges* = .34. According to Tukey HSD comparisons, for first-graders, high teacher expectations led to a 15.4-point IQ gain over normal teacher expectations (*p* < .001, 95% *CI of the difference* = 12.01 to 18.72). For sixth-graders, high teacher expectations led to only a 1.5-point IQ gain over normal expectations (*p* = .67).

This description of the interaction is effective because it is clear how the two simple effects differ. It may take several attempts for you to phrase your description of an interaction in a way that makes it clear to the reader. Do not expect your first attempt to be the best.

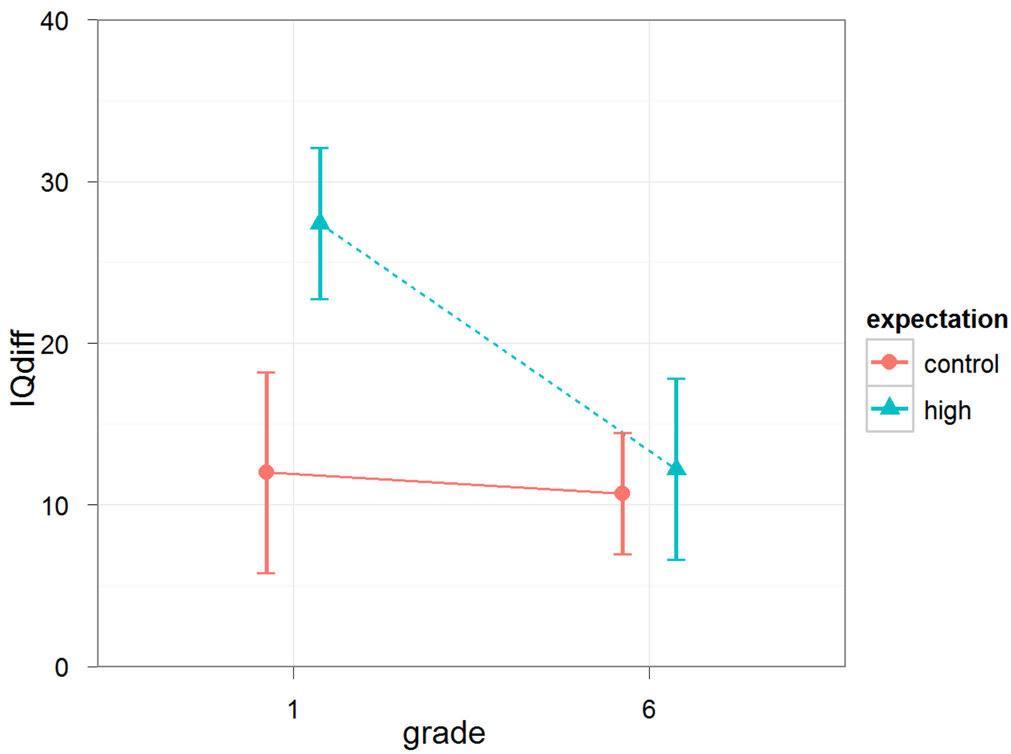
Interpreting Main Effects and Interactions through Figures

You can often get a quick idea of the pattern of results from a study by looking at a figure. It's important to remember that figures cannot tell you whether a pattern is significant – for that, you need the results of a statistical test.

Line graphs

Figure 5 provides a graphical representation of the mean IQ scores of the high- and normal-expectation first-grade and sixth-grade students. The dependent variable always goes on the y-axis and the two independent variables go along the x-axis and in the legend. Be sure to always label your x and y axes.

Figure 5. Effects of grade and teacher expectations on IQ scores.



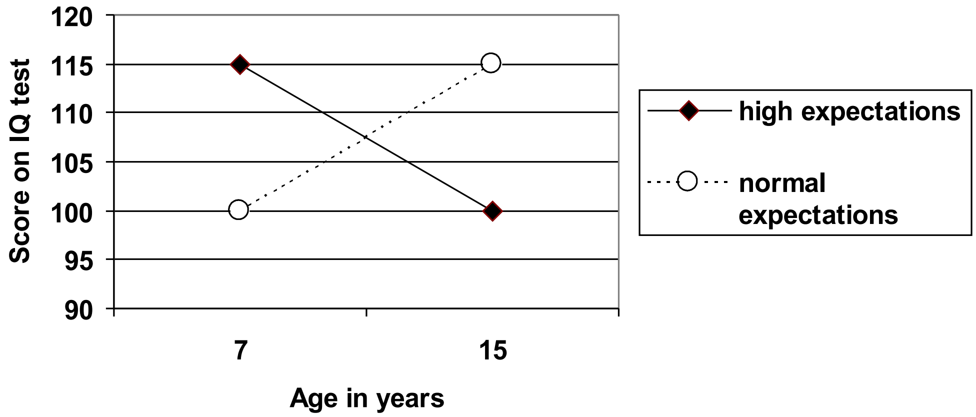
Interactions. The less parallel the lines are, the more likely there is to be a significant interaction. In Figure 5, we see that the lines are definitely *not* parallel, so we would expect an interaction.

Main effects. For the main effect of expectations, look to see whether the two pink circles (normal expectations) are, in general, higher or lower than the two blue triangles (high expectations) in Figure 5. Averaging across first- and sixth-graders, we would say that the blue triangles are generally higher than the open circles. Thus, we would expect a main effect of expectations such that high expectations lead to higher scores than normal expectations.

For the main effect of grade, look to see whether the average of the triangle and circle in the first-grade condition is higher or lower than the average of the triangle and circle in the sixth-grade condition. It looks like the first-grade average is higher than the sixth-grade average.

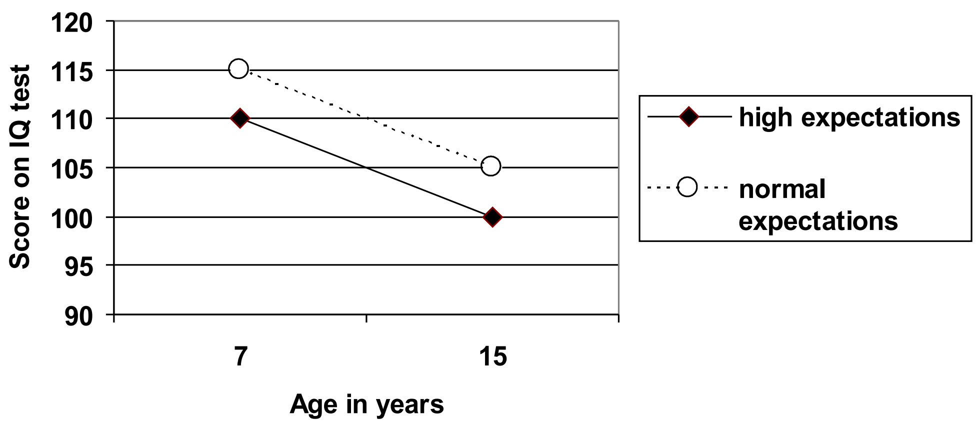
Below are some examples that depart from the dataset we have been examining to illustrate other possible effects you might see in a plot. Their labels (7-year-olds vs. 15-year-olds) are a bit different from the first-graders and sixth-graders we've been looking at, and the DV is in units of IQ score, not IQ gain.

Figure 6. A cross-over interaction.



Try to guess whether there is an interaction or any main effects in Figure 6. The lines are not parallel, so there is likely to be an interaction. Because the lines intersect, this type of interaction is sometimes called a cross-over interaction. It appears that the two black diamonds are not higher or lower than the two open circles, so there is no main effect of teacher expectation. It also appears that the average 7-year-old score is the same as the average 15-year-old score, so a main effect of age is unlikely. In summary, Figure 6 shows no main effect for age, no main effect for expectation, but a cross-over interaction.

Figure 7. Parallel lines.

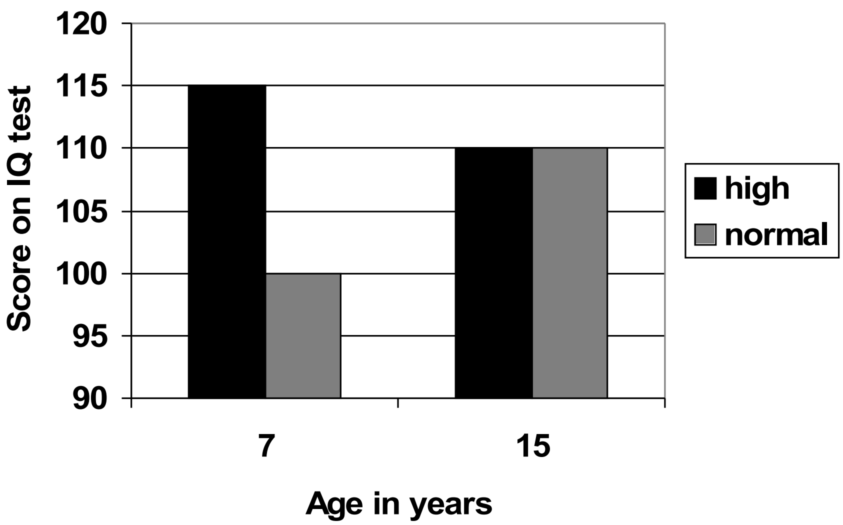


Try the same exercise with Figure 7. The lines are parallel, so there is no interaction. In general, the open circles are higher than the black diamonds, so there is likely to be a main effect of expectations such that normal expectations lead to higher IQ scores than high expectations. Also, the scores for 7-year-olds are higher than the scores for 15-year-olds, suggesting that there is a main effect for age. In summary, Figure 7 shows a main effect of age, a main effect of expectations, and no interaction.

As these examples demonstrate, main effects and interactions are independent of one another. You can have main effects without interactions, interactions without main effects, both, or neither.

Bar Graphs

Figure 8. Effects of age and teacher expectations on IQ scores.



Interpreting bar graphs for main effects and interactions is similar to line graphs, except that identifying interactions is harder and identifying main effects might be easier.

Interactions. To identify an interaction in a bar graph, look for **"a difference in differences."** First, look for the difference between high- and normal expectations for 7-year-olds: a difference of about 15 points. Now look at the difference between high and normal expectations for 15-year-olds: a difference of 0 points. The difference of 15 is *different* from the difference of 0. The differences are different. This is the mark of an interaction: the effect of one IV (teacher expectations) is different across different levels of another IV (age).

Main effects. Use the same logic as for the line graphs: are the black bars, in general, higher or lower than the lighter bars? Yes, the black bars are higher. Are the two bars for 7-year-olds higher or lower than the two bars for 15-year-olds? It looks like the bars for 15-year-olds might be slightly higher.

Conclusion

When you have more than one independent variable in a study, you have a *factorial design*. The *main effect* of each independent variable (or "factor") is the effect of that variable on the dependent variable, ignoring the effects of other independent variables. In addition to testing for the main effect of each factor, researchers are typically also interested in the *interaction* between the factors. A *statistical interaction* occurs when the effect of one independent variables changes depending on the level of another independent variable.

References

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behaviour Research Methods, 37* (3), 379-384.

Rosenthal, R., & Jacobson, L. (1966). Teachers' expectancies: Determinants of pupils' IQ gains. *Psychological Reports, 19,* 115-118.