

# Answers to Application Activities in Chapter 11

---

## 12.1.3 Application Activity: Identifying Between-Group and Within-Group Variables to Decide Between RM ANOVA and Factorial ANOVA Designs

1 Schön et al. (2008)

Dependent variable: Accuracy in correctly identifying words				
Independent variable(s):	1 Condition	No. of Levels:	3	Status of IV: Within-group <u>Between-group</u>
	2 L1	No. of Levels:	2	Within-group <u>Between-group</u>
For this research design, use: RM ANOVA <u>Factorial ANOVA</u>				

2 Erdener and Burnham (2005)

Dependent variable: Nonword accuracy				
Independent variable(s):	1 Condition	No. of Levels:	4	Status of IV: <u>Within-group</u> Between-group
	2 L1	No. of Levels:	2	Within-group <u>Between-group</u>
	3 Target lge	No. of Levels:	2	Within-group <u>Between-group</u>
For this research design, use: <u>RM ANOVA</u> Factorial ANOVA				

### 3 Larson-Hall (2004)

Dependent variable: Contrast perception scores				
Independent variable(s):	1 Contrasts	No. of	16	Status of IV
		Levels:	Within-group	Between-group
	2 Proficiency level	No. of	3	Within-group
		Levels:		Between-group
For this research design, use: RM ANOVA Factorial ANOVA				

### 4 Bitchener, Young, and Cameron (2005)

Dependent variable:	Percentage correct use			
Independent variable(s):	1 Treatment	No. of	3	Status of IV
		Levels:	Within-group <u>Between-group</u>	
	2 Writing assignment (Time)	No. of	4	<u>Within-group</u> Between-group
		Levels:		
	3 Type of error	No. of	3	<u>Within-group</u> Between-group
		Levels:		
For this research design, use:	<u>RM ANOVA</u>		Factorial ANOVA	

### 5 Ellis, Loewen, and Erlam (2006)

Dependent variable: Accuracy scores on whether past tense was correct or not				
Independent	1 Error	No. of	3	Status of IV
variable(s):	correction group	Levels:	Within-group	<u>Between-group</u>
	2 Type of task	No. of	3	<u>Within-group</u> Between-group
		Levels:		
	3 Time of test	No. of	2	<u>Within-group</u> Between-group
		Levels:		
For this research design, use:		<u>RM ANOVA</u>	Factorial ANOVA	

### 6 Flege, Schirru, and MacKay (2003)

Dependent variable: Listeners' judgments as to the accuracy of vowels				
Independent	1 Age group	No. of	2	Status of IV
variable(s):		Levels:	Within-group	<u>Between-group</u>
	2 Use of Italian	No. of	2	Within-group <u>Between-group</u>
		Levels:		
For this research design, use:		RM ANOVA	<u>Factorial ANOVA</u>	

## 7 Chrabaszcz & Gor (2014)

Dependent variable: Contrast discrimination scores				
Independent variable(s):	1 Consonant used in hard/soft pairing	No. of Levels:	3	Status of IV Within-group Between-group
	2 L1 group	No. of Levels:	2	Within-group Between-group
For this research design, use:		RM ANOVA Factorial ANOVA		

## 8 Xiao & Wong (2014)

Dependent variable: Anxiety score				
Independent variable(s):	1 Language area	No. of Levels:	4	Status of IV Within-group Between-group
	2 Type of learner	No. of Levels:	2	Within-group Between-group
For this research design, use:		RM ANOVA Factorial ANOVA		

## 9 Ayoun & Salaberry (2008)

Dependent variable:	Verb cloze task			
Independent variable(s):	1 Lexical aspect	No. of	4	Status of IV
	class	Levels:	Within-group	Between-group
For this research design, use:	RM ANOVA		Factorial ANOVA	

### 11.2.4 Application Activity for Changing Data from the Wide Form to the Long Form (R directions only as the SPSS files are in the correct form already and do not need to be converted)

#### 1 Lyster.Written.sav

This file is in the correct form to use for RM ANOVA in SPSS. It is in the wide form, with the one repeated-measure variable that we will be looking at (the Binary choice variable, with 3 different testing times) in 3 different columns, plus the variable of Cond, which specifies which of the 4 testing conditions (a between-groups variable) the participant received. For use with R, however, we need the data in the long form, so this answer explains how to create that file. Let's import the Lyster.Written.sav file into R and call it LysterBinary.

```
names(LysterBinary)
```

```
library(reshape2) #open the library for the melt() command
```

```
LysterB<-melt(LysterBinary[c(1:5)], #choose columns 1-5
```

```
id.vars=c("participant", "cond"), #ID between-groups columns
```

```
value.name="binaryscore",      #give name to stacked score from the 3 time periods  
variable.name="time") #name the column that labels the index for time period
```

```
levels(LysterB$time)=c("pre", "immediate post", "delayed post") #add names  
str(LysterB)
```

To save this file now, in R Commander make **LysterB** your active dataset, choose DATA > ACTIVE DATASET > EXPORT ACTIVE DATASET, choose “Comma” as your field separator, then press OK. Navigate to wherever you want to save the file and save it. The R command for doing this was:

```
write.table(LysterB, "C:/Users/LarsonHall/Desktop/LysterB.csv", sep=",",  
col.names=TRUE, row.names=TRUE, quote=TRUE, na="NA")
```

## 2 Cole1927.sav

This file is in the correct form to use for RM ANOVA in SPSS in the wide form, but needs to be put into the long form for use with R. There are two repeated measures, time with two levels and measure type with three levels, meaning that there are 6 columns of the scores. To use with R, change the data to the long form with only one column of scores, but with two index columns for the repeated measure, one for testing time and another for measure type. We also need one index for the independent variable of teaching group (with two levels) and a participant index. Import the Cole1927.sav file into R and call it **cole1927**.

```
names(cole1927)
```

```
length(cole1927$group) #find out how many rows in this dataset
```

```
library(reshape2) #open the library for the melt( ) command
```

```
cole1927<-cbind(cole1927, participant=factor(rep(c(1:30)))) # make the participant index
```

```
#from the length( ) command above I learned there were 30 rows
```

```
names(cole1927) #verify the participant variable is now at the end of the dataset
```

```
ColeLong<-melt(cole1927[c(1:7,14)],
```

```
id.vars=c("participant", "group"), #ID between-groups columns
```

```
value.name="score") #give name to stacked score from the 6 columns
```

```
> str(ColeLong)
'data.frame': 180 obs. of 4 variables:
 $ participant: Factor w/ 30 levels "1","2","3","4",...: 1 2 3 4 5 6 7 8 9 10 ...
 $ group      : Factor w/ 2 levels "Composition",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ variable   : Factor w/ 6 levels "time1_vocab_percent",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ score      : num 76 73.3 62.7 74.7 70.7 ...
```

We need the one index variable for the repeated measure (called **variable** now) to be split into two indexes, one for time and one for measuretype.

```
ColeLong<-cbind(ColeLong, colsplit(ColeLong$variable, "_", names=c("time",
"TestType")))
```

```
> str(ColeLong)
'data.frame': 180 obs. of 6 variables:
 $ participant: Factor w/ 30 levels "1","2","3","4",...: 1 2 3 4 5 6 7 8 9 10 ...
 $ group      : Factor w/ 2 levels "Composition",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ variable   : Factor w/ 6 levels "time1_vocab_percent",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ score      : num 76 73.3 62.7 74.7 70.7 ...
 $ time       : chr "time1" "time1" "time1" "time1" ...
 $ TestType   : chr "vocab percent" "vocab percent" "vocab percent" "vocab percent"
```

There is now one index for participant (**participant**), one for group (**group**), one for testing time (**time**) and one for measure type (**TestType**). The dependent variable (**score**) is stacked 6 times ( $6 \times 30 = 180$ ), appropriately. The only problems are that the variable called **variable** is not necessary, the variable **TestType** has “percent” appended to it but doesn’t need it, and then it looks like the variable called time only has one label. In that case, the first three variable should have been called time 1 but the second three should have been called time 2. First I’ll get rid of the extraneous variable:

```
ColeLong$variable <- NULL
```

Now let’s check the levels of the **time** variable:

```
levels(ColeLong$time)
```

It says the answer is NULL, which is worrying until I remember that for my new index variables of **time** and **TestType** they are not factors yet (see the structure, printed above)! I need to make them into factors.



```
ColeLong$time<-as.factor(ColeLong$time)
```

```
ColeLong$TestType<-as.factor(ColeLong$TestType)
```

Now if I look specifically at the levels of these factors I see that the levels have been appropriately resolved into the correct number:

```
> levels(ColeLong$time)
[1] "time1" "time2"
> levels(ColeLong$TestType)
[1] "grammar_percent" "reading_percent" "vocab_percent"
```

To save this file now, in R Commander make **ColeLong** your active dataset, choose DATA > ACTIVE DATASET > EXPORT ACTIVE DATASET, choose “Comma” as your field separator, then press OK. Navigate to wherever you want to save the file and save it. The R command for doing this was:

```
write.table(ColeLong, "C:/Users/LarsonHall/Desktop/ColeLong.csv", sep=",",
col.names=TRUE, row.names=TRUE, quote=TRUE, na="NA")
```

### 3 LarsonHall2004.sav

This file is in the correct form to use for RM ANOVA in SPSS in the wide form but needs to be put into the long form for use with R. There is one repeated measure, which is phonological contrast with three levels. Import the LarsonHall2004.sav file into R and call it **lh2004**.

```
names(lh2004)
```

```
library(reshape2) #open the library for the melt( ) command
```

Create a participant variable:

```
length(lh2004$level)
```

```
lh2004<-cbind(lh2004, participant=factor(rep(c(1:41)))) # make the participant index
```

#from the `length( )` command above I learned there were 41 rows

```
names(lh2004) #verify the participant variable is now at the end of the dataset
```

```
lh2004Long<-melt(lh2004[c(1, 5,7,11, 19)],
```

```
id.vars=c("participant", "level"),
```

```
value.name="score", #give name to stacked score from the 3 contrasts
```

```
variable.name="contrast") #name the column that labels the index for the
```

#stacked scores for the contrasts

```
str(lh2004Long)
```

To save this file now, in R Commander make `lh2004Long` your active dataset, choose DATA > ACTIVE DATASET > EXPORT ACTIVE DATASET, choose “Comma” as your field separator, then press OK. Navigate to wherever you want to save the file and save it. The R command for doing this was:

```
write.table(lh2004Long, "C:/Users/LarsonHall/Desktop/lh2004Long.csv",  
sep=",", col.names=TRUE, row.names=TRUE, quote=TRUE, na="NA")
```

#### 4 Erdener&Burnham2005.sav

This file is in the correct form to use for RM ANOVA in SPSS in the wide form, but needs to be put into the long form for use with R. There is one repeated measure, which is condition with four levels. Import the Erdener&Burnham2005.sav file into R and call it `erdener`.

```
names(erdener)
```

```
length(erdener$1) #find out how many rows in this dataset
```

```
library(reshape2) #open the library for the melt( ) command
```

```
erdener<-cbind(erdener, participant=factor(rep(c(1:64)))) # make the participant index
```

#from the `length( )` command above I learned there were 67 rows

```
names(erdener) #verify the participant variable is now at the end of the dataset
```

```
ErdenerLong<-melt(erdener, #use all of the variables in this dataset
```

```
id.vars=c("participant", "1"), #ID between-groups columns
```

```
value.name="score") #give name to stacked score from the 6 columns
```

```
> str(ErdenerLong)
'data.frame': 536 obs. of 4 variables:
 $ participant: Factor w/ 67 levels "1","2","3","4",...: 1 2 3 4 5 6
 $ l1         : Factor w/ 2 levels "Turkish","English": 1 1 1 1 1 1
 $ variable   : Factor w/ 8 levels "sp_audonly","sp_av",...: 1 1 1 1
 $ score      : num 0 0 4 6 5 2 5 4 3 3 ...
```

We need the one index variable for the repeated measure (called **variable** now) to be split into two indexes, one for target language and one for condition.

```
ErdenerLong<-cbind(ErdenerLong, colsplit(ErdenerLong$variable, "_", names=c("TL",
"Condition")))
```

```
> str(ErdenerLong)
'data.frame': 536 obs. of 6 variables:
 $ participant: Factor w/ 67 levels "1","2","3","4",...: 1 2 3 4
 $ l1         : Factor w/ 2 levels "Turkish","English": 1 1 1 1
 $ variable    : Factor w/ 8 levels "sp_audonly","sp_av",...: 1 1
 $ score       : num 0 0 4 6 5 2 5 4 3 3 ...
 $ TL         : chr "sp" "sp" "sp" "sp" ...
 $ Condition   : chr "audonly" "audonly" "audonly" "audonly" ...
```

There is now one index for participant (**participant**), one for group (**L1**), one for target language (**TL**) and one for condition (**condition**). The dependent variable (**score**) is stacked 8 times ( $8 \times 67=536$ ), appropriately. Let's get rid of the extraneous variable:

```
ErdenerLong$variable <- NULL
```

And because the **str()** command shows that TL and Condition are not factors yet, make them factors:

```
ErdenerLong$TL<-as.factor(ErdenerLong$TL)
```

```
ErdenerLong$Condition<-as.factor(ErdenerLong$Condition)
```

Now if I look specifically at the levels of these factors I see that the levels have been appropriately resolved into the correct number:

```
levels(ErdenerLong$TL)
```

```
[1] "irish" "sp"
```

```
levels(ErdenerLong$Condition)
```

```
[1] "audonly" "audorth" "av" "avorth"
```

To save this file now, in R Commander make **ErdenerLong** your active dataset, choose DATA > ACTIVE DATASET > EXPORT ACTIVE DATASET, choose “Comma” as your field separator, then press OK. Navigate to wherever you want to save the file and save it. The R command for doing this was:

```
write.table(ErdenerLong, "C:/Users/LarsonHall/Desktop/ErdenerLong.csv",
```

```
sep="," , col.names=TRUE, row.names=TRUE, quote=TRUE, na="NA")
```

## 5 Toth2008Prod.sav

This file is in the correct form to use for RM ANOVA in SPSS in the wide form but needs to be put into the long form for use with R. There is one repeated measure, which is time with three levels. Import the Toth2008Prod.sav file into R and call it `tothProd`.

```
names(tothProd)
```

```
library(reshape2) #open the library for the melt( ) command
```

The variable subject is a participant variable so we do not need to create one ourselves.

```
tothProdLong<-melt(tothProd,
```

```
id.vars=c("subject", "group"),
```

```
value.name="score", #give name to stacked score from the 3 contrasts
```

```
variable.name="time") #name the column that labels the index for the
```

```
#stacked scores for the contrasts
```

```
str(tothProdLong)
```

To save this file now, in R Commander make `tothProdLong` your active dataset, choose DATA > ACTIVE DATASET > EXPORT ACTIVE DATASET, choose “Comma” as your field separator, then press OK. Navigate to wherever you want to save the file and save it. The R command for doing this was:

```
write.table(tothProdLong, "C:/Users/LarsonHall/Desktop/tothProdLong.csv",  
sep=",", col.names=TRUE, row.names=TRUE, quote=TRUE, na="NA")
```

### 11.3.5 Application Activity with Parallel Coordinate Plots (both SPSS and R instructions given)

#### 1 Murphy (2004)

For SPSS, use the Murphy.RepeatedMeasures.sav file. For R use the **murphy.wide** file (if you have followed along with me in the text you'll already have this imported, but if not, import the Murphy.RepeatedMeasures.sav and name it **murphy.wide**).

#### SPSS Instructions:

Use the menu sequence GRAPHS > GRAPHBOARD TEMPLATE CHOOSER. When the Template Chooser opens, click on the six repeated variables in the file: REGULAR PROTOTYPICAL, REGULAR INTERMEDIATE, REGULAR DISTANT, IRREGULAR PROTOTYPICAL, IRREGULAR INTERMEDIATE, and IRREGULAR DISTANT (do this by holding down the CTRL button while clicking with the mouse). SPSS recommends a number of types of plots for this type of data, so choose the graph labeled “Parallel.” Press OK.

Actually, you'll see that with the Murphy.RepeatedMeasures.sav file, you will *not* get a choice of “Parallel” graph. This is because the six repeated measures variables in the dataset have been defined as ordinal measurements instead of what SPSS calls “scale” measurements (and which I call “interval” or “continuous” measurements). In order to get a parallel coordinate plot, change the variable type to ordinal. Do this by going to the “Variable View” tab, then click on one of the

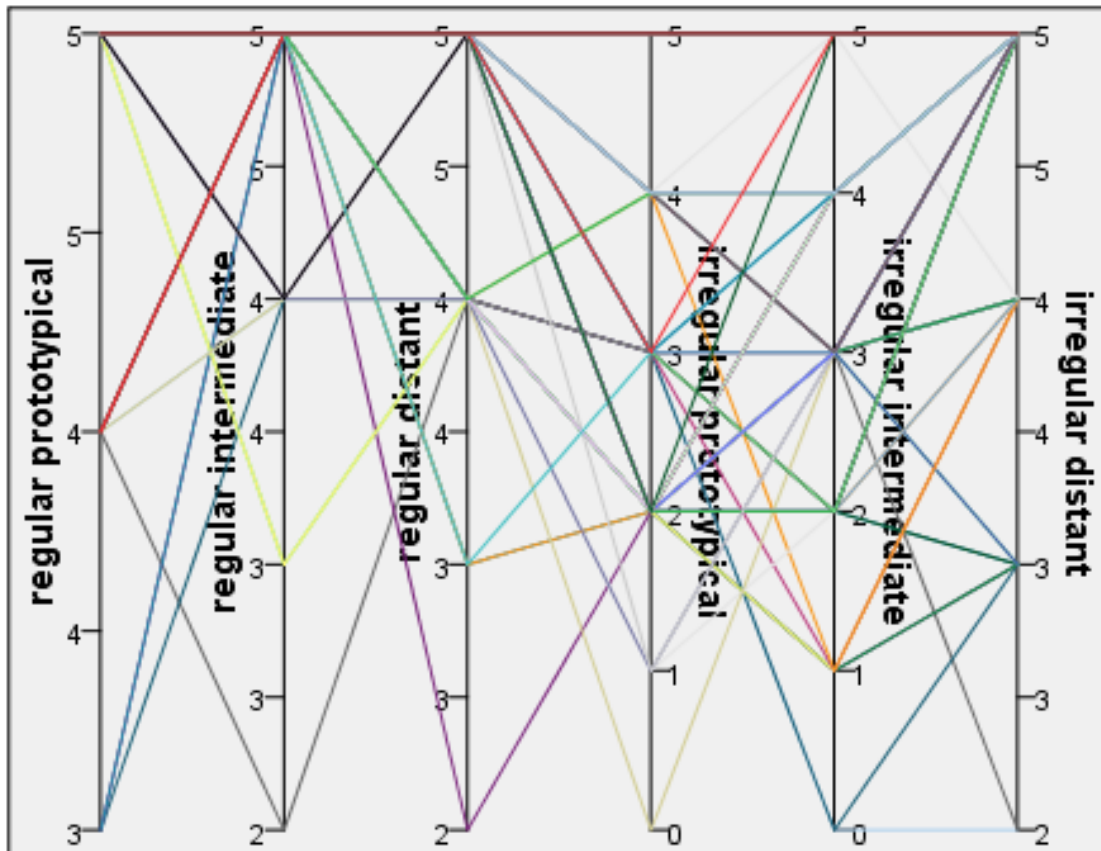
six variables. Next, choose the menu choice DATA > DEFINE VARIABLE PROPERTIES. Move all 6 of the repeated variables to the right (you can highlight the first, then hold down the Shift button and highlight the last to highlight all 6) and move the variables to the right, to the “Variables to Scan” box. Press CONTINUE. In the following dialogue box, called “Define Variable Properties,” click on each variable in turn and change the “Measurement Level” to “Scale.” You can either do this one by one, or just change the first one then press the “To Other Variables” button in the “Copy Properties” area, and then click on the 5 other variables when the “Apply Labels and Level to” box pops up. Press OK when you are finished. You should now be able to follow the instructions above.

The graph that results doesn’t look so nice, and one reason is that that there were only 5 points in each measurement and it turns out that there were only a couple of patterns for each of the three groups. Another problem with the graph is that the scales for each of the 6 variables are not lined up to compare against each other (in other words, they don’t all start at 0 and end at 5). We could also split the data by groups, which you can do by going to DATA > SPLIT FILE ticking the “Organize output by groups” button and then putting the SUBJECTS variable into the “Groups Based On” box. Press OK. Run the graphic again, and you’ll get 3 graphs where the first is Group 1 (NS children), second is Group 2 (NS adults) and third is Group 3 (NNS adults). For each of these you can change the minimum and maximum scale (double-click on the graph to open the GRAPHBOARD EDITOR; click on the vertical line that has the scale and at the bottom of the GRAPHBOARD EDITOR tabs will appear. Click on the tab called SCALE. Here you can set the “Min” and “Max” values for your axes. I don’t know of any way to jitter lines so the view can



get a sense of how many participants chose the same number though, so it looks like a parallel coordinate plot is really not the best choice for displaying this data.

Here is the original graph (before splitting the file):

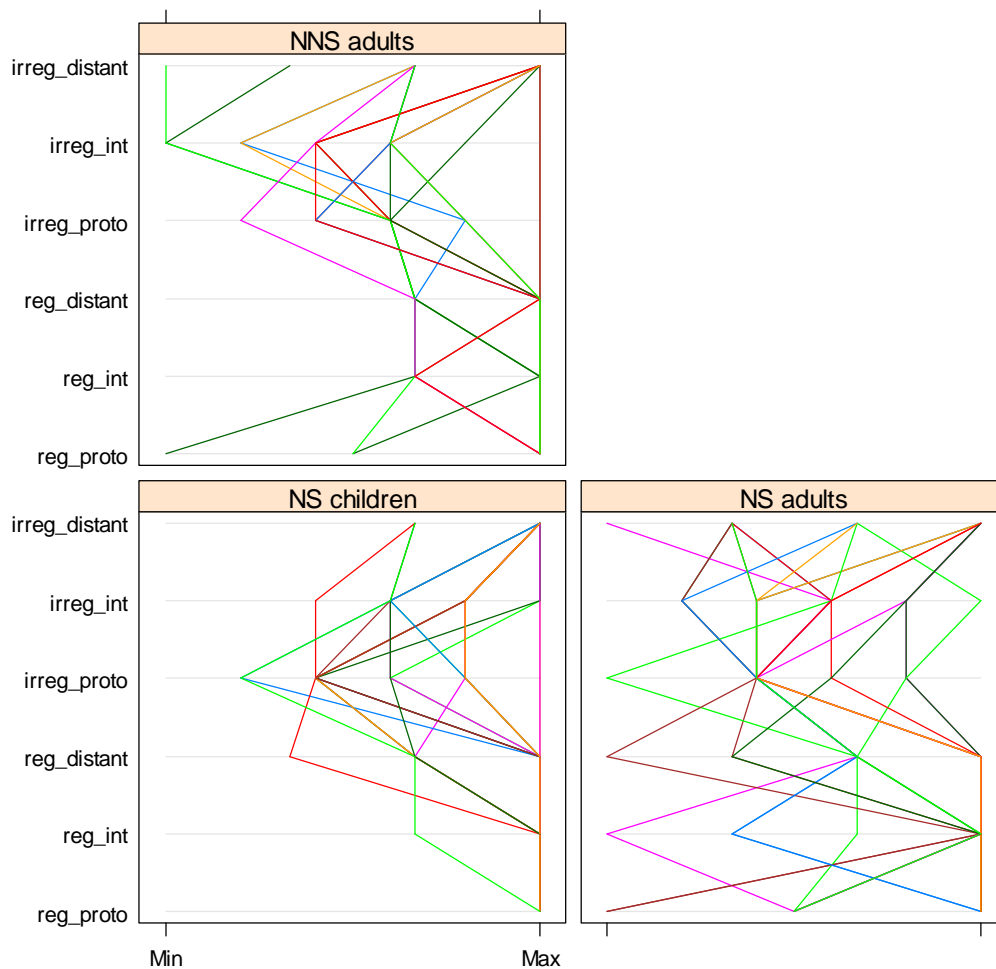


**R Instructions:**

```
library(lattice)
```

```
parallelplot(~murphy.wide[2:7] | murphy.wide$group, data=murphy.wide)
```

Here is what the graphic looks like without any further modification:



The reason that there are only a couple of lines for each group, in spite of there being 20 different participants in each group, is that there were only 5 points in each measurement and it turns out that there were only a couple of patterns for each of the three groups. In other words, a large number of persons picked exactly the same number in the limited 5-point scale, but we can only see one point there so we can't tell how many people chose that point. I don't know of any way to jitter lines so the view can get a sense of how many participants chose the same number

though, so it looks like a parallel coordinate plot is really not the best choice for displaying this data.

## 2 Lyster (2004)

For SPSS, use the Lyster.Written.sav file. For R import this file as `lyster.wide`.

### **SPSS Instructions:**

First, split the file (DATA > SPLIT FILES, then move COND over). Then choose GRAPHS > GRAPHBOARD TEMPLATE CHOOSER and choose PREBINARYCHOICE, POST1BINARYCHOICE, and POST2BINARYCHOICE variables (do this by holding down the CTRL button while clicking with the mouse). Choose the graph labeled “Parallel.” Press OK.

You’ll have 4 different graphs, but none have titles, nor are their scales aligned. I recommend you double-click on the graph to open the GRAPHBOARD EDITOR; click on the vertical line that has the scale and at the bottom of the GRAPHBOARD EDITOR tabs will appear. Click on the tab called SCALE. Here you can set the “Min” and “Max” values for your axes, and set them to all run from the same minimum to maximum. I also recommend copying each edited graph to a different program, like the Paint program, so that you can put titles identifying each group on them (look at the “Values” for COND in the Variable View tab to see the order of the groups; they will be printed in the same order as the Value Labels).

The FFIPrompt group again seems to have the most participants who improve strongly, although there are a few exceptions to this rule. There is a trend toward some improvement for FFIRecast, and for the other two conditions it looks fairly random which way the lines go.

### **R Instructions:**

```
library(lattice)
```

```
parallelplot(~lyster.wide[3:5] | lyster.wide$cond, data=lyster.wide)
```

The FFIPrompt group again seems to have the most participants who improve strongly, although there are a few exceptions to this rule. There is a trend toward some improvement for FFIRecast, and for the other two conditions it looks fairly random which way the lines go.

### **3 Ayoun & Salaberry (2008)**

For SPSS, use the AyounSalaberryCloze.sav file. For R import this file as **ayoun**.

### **SPSS Instructions:**

Choose **GRAPHS > GRAPHBOARD TEMPLATE CHOOSER** and choose the **TELICS, ACTIVITIES, and STATES** variables (do this by holding down the **CTRL** button while clicking with the mouse).

You'll find that you get no graph choice of "Parallel," and you might notice this is because your variables are labeled as nominal (categorical) variables (represented with an icon that has three colored circles). Change the variables to what SPSS calls "Scale" variables (Interval or continuous variables) by going to the "Variable View" tab, then clicking on one of the six

variables. Next, choose the menu choice DATA > DEFINE VARIABLE PROPERTIES. Move all 3 variables to the right, to the “Variables to Scan” box. Press CONTINUE. In the following dialogue box, called “Define Variable Properties,” the previous measurement as “Nominal” will have already all changed to “Scale” (with a ruler icon) because SPSS scanned the data and found it would be best represented by an interval measurement. However, to confirm the changes you will have to do something to get the OK button to appear, so just click on the “Measurement Level” and choose “Scale” once. Then press OK. Now you can run the graph above.

Again, this graphic will be misleading if you leave the scale as it is, as in the original graphic (reprinted below) it looks like participants scored much higher on Telics than States, but a look at the scale shows they actually scored more highly on States! Change the scale as described in the answers for Exercise #1 or #2 of this section.

The strong trend in the data is for most participants to score more highly on Telics and States than they do on Activities. There seems to be a slight trend to score a little bit higher on Telics than on States as well.

**R Instructions:**

```
library(lattice)
```

```
parallelplot(~ayoun)
```

The strong trend in the data is for most participants to score more highly on Telics and States than they do on Activities. There seems to be a slight trend to score a little bit higher on Telics than on States as well.

#### 4 Cole (1927)

For SPSS, use the Cole1927.sav file. For R import this file as **cole1927**.

#### **SPSS Instructions:**

First, split the file by teaching group (DATA > SPLIT FILES, then move GROUP over). Then choose GRAPHS > GRAPHBOARD TEMPLATE CHOOSER and choose the 6 variables that have “TIME\_” at the beginning of their name (do this by holding down the CTRL button while clicking with the mouse). Choose the graph labeled “Parallel.” Press OK.

There are two graphs, the first for the “Composition” group and the second for the “Grammar Translation” group. These graphs are so messy that I think it might be better to only look at the graphs with the Time 1 and Time 2 variables for each test type separately.

Do this by going back to the Graphborad Template Choose and choosing only, for example, TIME1\_VOCAB\_PERCENT and TIME2\_VOCAB\_PERCENT. This graph is much easier to understand! Repeat with the other variables. If I wanted to print these graphs in my research report, I’d copy them into Paint and label them nicely and gather them together. I’d also want to change the scale of the axes to be the same as described in the answers for Exercise #1 or #2 of this section.

As for trends, there are a lot of straight lines, indicating that many participants did not gain in scores from Time 1 to Time 2. The language measure with the most rightward slant (and thus the most increase from Time 1 to Time 2) looks to be grammar, with the Grammar Translation group especially strong here. The Composition group has some good increases but several lines also show loss from Time 1 to Time 2 as well. For vocabulary the Composition group seems to have less variation (the lines are clumped together more strongly) than for the Grammar Translation group. For reading most lines are straight.

### **R Instructions:**

```
library(lattice)
```

```
parallelplot(~cole1927[2:7] | cole1927$group, data=cole1927)
```

These graphs are so messy that I think it might be better to only look at the graphs with the Time 1 and Time 2 variables for each test type separately. I used this code:

```
parallelplot(~cole1927[c(2,5)] | cole1927$group, data=cole1927)
```

```
parallelplot(~cole1927[c(3,6)] | cole1927$group, data=cole1927)
```

```
parallelplot(~cole1927[c(4,7)] | cole1927$group, data=cole1927)
```

Note: In some commands to pick out two different columns you can just list them, like

```
cole1927[2, 5]
```

but doing this for the `parallelplot( )` command returns an error so I tried the `c( )`

command, which lets you list things, and it worked.

Note: I also tried to use the parameter command to put these graphs into a more compact form with all of them in the same window (`par(mfrow(3, 1))`), but that didn't work so what I would do to make this look better would be to copy each graphic (FILE > COPY TO CLIPBOARD > (pick Bitmap or Metafile) and put them together in another program like the Paint program).

As for trends, there are a lot of straight lines, indicating that many participants did not gain in scores from Time 1 to Time 2. The language measure with the most rightward slant (and thus the most increase from Time 1 to Time 2) looks to be grammar, with the Grammar Translation group especially strong here. The Composition group has some good increases but several lines also show loss from Time 1 to Time 2 as well. For vocabulary the Composition group seems to have less variation (the lines are clumped together more strongly) than for the Grammar Translation group. For reading most lines are straight.

## 5 Toth (2008)Prod.sav

For SPSS, use the Toth2008Prod.sav file. For R import this file as `tothProd`.

### SPSS Instructions:

First, split the file by group (DATA > SPLIT FILES, then move GROUP over). Then choose GRAPHS > GRAPHBOARD TEMPLATE CHOOSER and choose the PRODPRETEST, PRODPOSTTEST, and DELAYEDPOSTTEST variables (do this by holding down the CTRL button while clicking with the mouse). Choose the graph labeled "Parallel." Press OK. What you might immediately notice is that the scale is very strange! The scale has several ticks on it that all say "0" or "1"! If you look at the file you'll see that "0"s and "1"s are the only numbers in the file. Because there are a



number of “0”s this probably means there are smaller divisions. So I went to the “Variable View” tab and clicked on the cell under “Decimals” for the PRODPRETEST. When I changed it to 2 decimals, I was able to see that there were indeed values like .25 and .75. So I ran the graphic command again after changing the decimals shown, and now the scales made much more sense. To make this a better graphic, change the scale as described in the answers for Exercise #1 or #2 of this section, and put titles on the charts to label which groups produced which graphics.

The plots show that there was much more variation in scores for the Teacher-led and Learner-led groups than for the Control group. Comparing the scores with the graphic, I see that there are quite a lot of the controls who followed the pattern of getting a zero on every test, but from the lines in the Control group graphic, I can’t tell how many people got the same scores. Like the Murphy dataset in Exercise #1, this graphic would benefit from some jitter of the lines, but there’s no way to get that, so the parallel coordinate plot might not be the best graphic for this data. Beyond that, the experimental groups show wide variation. It looks like a lot of the participants started with nothing in the pretest, but then varied widely in the posttest and delayed posttest, with some gaining a lot of points in the posttest but then dropping back to zero in the delayed posttest. These graphs would make me think that there’s so much variation in the data that there are no real good patterns to it!

**R Instructions:**

```
library(lattice)
```

```
parallelplot(~tothProd[3:5] | tothProd$group, data=tothProd)
```

The plots show that there was much more variation in scores for the Teacher-led and Learner-led groups than for the Control group. Comparing the scores with the graphic, I see that there are quite a lot of the controls who followed the pattern of getting a zero on every test, but from the lines in the Control group graphic, I can't tell how many people got the same scores. Like the Murphy dataset in Exercise #1, this graphic would benefit from some jitter of the lines, but there's no way to get that, so the parallel coordinate plot might not be the best graphic for this data. Beyond that, the experimental groups show wide variation. It looks like a lot of the participants started with nothing in the pretest, but then varied widely in the posttest and delayed posttest, with some gaining a lot of points in the posttest but then dropping back to zero in the delayed posttest. These graphs would make me think that there's so much variation in the data that there are no real good patterns to it!

### **11.5.4 Application Activities with least-squares RM ANOVA (SPSS and R answers given)**

#### **1 Lyster (2004) Binary Choice data**

Use the Lyster.Written.sav file. For R, use the **LysterB** dataset you created in Section 11.2.4.

#### **SPSS Instructions:**

To conduct an RM ANOVA, use ANALYZE > GENERAL LINEAR MODEL > REPEATED MEASURES (remember, if you have previously split this file, you should go in and un-split it! The Data Editor has a little message at the bottom right that will say something like "Split by Cond" if your file is split. Go to DATA > SPLIT FILE and choose the "Analyze all cases" radio button. There is only one within-group factor, which is the different times the test was administered. Thus, go to the "Within-Subject Factor Name" box and label your variable TIME and specify it

has three levels in the “Number of Levels” box. ADD it, then click DEFINE. Move the 3 variables for the task completion test (PREBINARY, POST1BINARY, POST2BINARY) into the “Within-subjects” box. Put the COND variable in the “Between-subjects” box. Open the MODEL button and change to Type II SS. Create some interaction plots in the PLOTS button. Call for post-hocs on COND (use LSD and Games-Howell) through the POSTHOC button. Open the SAVE button and tick “Cook’s Distance” and “Unstandardized”. Open the OPTIONS button and tick the boxes for “Descriptive statistics,” “Estimates of effect size,” “Residual SSCP matrix,” “Spread vs. Level plot,” and “Residual plot”. Click CONTINUE then OK.

### **R Instructions:**

```
library(ez)
```

```
lysterModel1<-ezANOVA(data=LysterB, dv=.(binaryscore), wid=.(participant),  
within=.(time), between=.(cond)) #main RM-ANOVA command
```

I get the following warning because groups are unequal, but I am OK with it.

```
Warning: Data is unbalanced (unequal N per group). Make sure you specified a well-  
considered value for the type argument to ezANOVA().
```

```
print(lysterModel1)
```

```
(lysterStats<-ezStats(data=LysterB, dv=.(binaryscore), wid=.(participant),  
within=.(time), between=.(cond))) # descriptive statistics
```

## Results:

*Descriptive statistics.* First off, notice that the number of participants in this study is large, with  $N=180$  total and about 40 participants or more in each group. Looking at the descriptive statistics, we might notice that for the pre-task scores are roughly equal, but in the post-task (both immediate and delayed), the FFIPrompt group has the highest scores. However, all groups have improved relative to the Comparison group.

*Sphericity.* SPSS says that Mauchly's test of sphericity has a p-value of  $p=.000$ , which means that we have violated the sphericity assumption. R says that sphericity is violated for both the Time variable and for the interaction between Time and Condition (in both cases  $p<.0005$ ). Use the Greenhouse-Geisser or Huynh-Feldt correction when reporting interactions and main effects.

*Interactions and main effects.* In SPSS, look at the "Test of Within-Subjects Effects" box, and in R, look at the lines under \$ANOVA directly under the `print(lysterModel1)` line. Results here are the same whether from SPSS or R.

Time\*Condition:  $F_{6,352}=14.39$ ,  $p<.0005$  (Greenhouse Geisser correction)

Time:  $F_{2,352}=70.54$ ,  $p<.0005$  (Greenhouse Geisser correction)

Condition:  $F_{3,176}=15.50$ ,  $p<.0005$  (in SPSS this information is found later on in the output from the "Tests of Between-Subjects Effects" box).

The interaction and both main effects are statistical. With a statistical interaction, we don't really care about the main effects being statistical, so we'll need to find out more information about how Time varies given the Condition.

However, one point to think about is that we'd like for all the groups to be equal on the pretest, so the fact that the groups are different at different time periods may make us uneasy. An easy way to check on this would be to run a one-way ANOVA on the pre-test data (even though not finding a difference between groups is not necessarily evidence that there *is* no difference between groups!). Doing this in SPSS I called for a one-way ANOVA on the PREBINARY variable by using ANALYZE > COMPARE MEANS > ONE-WAY ANOVA. I put PREBINARY in the "Dependent List" box and COND in the "Factor" box. Not remembering whether variances were equal, when I opened the POSTHOC button I just ticked the Games-Howell post-hoc and then OK. Doing this in R Commander, I made **lyster.wide** the active dataset, then opened up STATISTICS > MEANS > ONE-WAY ANALYSIS OF VARIANCE. I chose **cond** for "Groups" and prebinary for the "Response Variable." I ticked the "Pairwise comparison of means" box, then ran the analysis.

The omnibus test had a result of  $F_{3,176}=2.46$ ,  $p=0.065$ . This doesn't make us feel very safe! In R I got CIs for the multiple comparisons between groups and one, the one between the Comparison and the FFI only groups did not pass through zero, although it was quite close to it [-7.18, -0.08]. Before I would assert that the groups were equivalent in the pretest I would try out my strategy for testing for equivalence between groups in a pretest that I outlined in Section 9.4.2). If I felt I couldn't assert the equivalence of my groups in the pretest, I might decide to run an RM

ANCOVA as a check on my results later, using the pre-test as a control variable (information about how to run ANCOVAs can be found in the online document "Chp12\_ANCOVA").

*Residual SSCP Matrix (SPSS only).* Look at the second row, the Covariance; here numbers along the diagonal should all be similar. Off-diagonal elements should be equal. This is basically true. Therefore, we have conflicted evidence—the Mauchley’s test indicates that sphericity does not hold while the Residual SSCP matrix indicates that sphericity does hold.

*Post-hoc tests.* The post-hocs given in the output tell us which groups differed from each other, but it aggregates the data across all time periods, and since we know the interaction is statistical, this isn’t very informative. However, I will ask for further analysis of this data in Section 11.6.3 so I will stop here for now.

*Assumptions (SPSS only).* The spread vs. level plot does not have many points, but they seem to be randomly distributed, upholding the assumption of homogeneity of variances in the data. In the residual plots examine the Std. Residual vs. Predicted table. The data are parallel lines, which do not appear randomly distributed, a result that would violate the assumption of homogeneity of variances. Look at histograms of the 6 columns with RES\_ at the end of the dataset (GRAPHS > LEAGCY DIALOGS > HISTOGRAMS). These don’t look too abnormal, so we might not worry too much that the data are not normally distributed.

For the Cook’s distance numbers that were appended to the end of the dataset, we want to make sure none of these were over 1. The dataset is large so perhaps the fastest way to do this is to ask

for descriptive stats for this: ANALYZE > DESCRIPTIVE STATISTICS > DESCRIPTIVES. Move over all the Cook's Distance variables. The mean score was .01 with no number being over 1, so there are no problems with outliers in this dataset.

## 2 Cole (1927)

For SPSS use the Cole1927.sav file. For R, use the **ColeLong** dataset you created in Section 11.2.4.

### **SPSS Instructions:**

To conduct an RM ANOVA, use ANALYZE > GENERAL LINEAR MODEL > REPEATED MEASURES (remember, if you have previously split this file, you should go in and un-split it! The Data Editor has a little message at the bottom right that will say something like “Split by . . .” if your file is split. Go to DATA > SPLIT FILE and choose the “Analyze all cases” radio button). There are two within-group factors for this dataset—the different times the test was administered (pre and post) and the three different types of tests/measures. Thus go to the “Within-Subject Factor Name” box and label your first variable TIME and specify it has two levels in the “Number of Levels” box. ADD it. Go back to the “Within-Subject Factor Name” box and enter the next variable of MEASURE, specify that it has 3 levels, click the ADD button, then click DEFINE. Move the 6 variables (TIME1\_VOCAB, TIME1\_GRAMMAR, TIME1\_READING, TIME2\_VOCAB, TIME2\_GRAMMAR, TIME2\_READING) into the “Within-subjects” box. Put the GROUP variable in the “Between-subjects” box. Open the MODEL button and change to Type II SS. Create some interaction plots in the PLOTS button. There is no need to call for post-hocs on GROUP because there are only two groups. Open the SAVE button and tick “Cook's Distance” and

“Unstandardized.” Open the OPTIONS button and tick the boxes for “Descriptive statistics”, “Estimates of effect size,” “Residual SSCP matrix,” “Spread vs. Level plot,” and “Residual plot.” Click CONTINUE then OK.

### **R Instructions:**

```
library(ez)  
  
ColeLongModel1<-ezANOVA(data=ColeLong, dv=.(score), wid=.(participant),  
within=.(time, TestType), between=.(group)) #main RM-ANOVA command  
  
print(ColeLongModel1)  
  
(ColeStats<-ezStats(data=ColeLong, dv=.(score), wid=.(participant),  
within=.(time, TestType), between=.(group))) #descriptive statistics
```

### **Results:**

*Descriptive statistics.* First off, notice that the number of participants in this study is small, with N=30 total and only 15 participants, although having repeated measures makes the results stronger. But unless effect sizes are fairly large, it is unlikely that results will be statistical. Looking at the descriptive statistics, we might notice that in comparing the groups to each other, they seem to be very similar in scores except for maybe the Time 2 vocab and Time 2 reading (and even here they are only about 5 points different out of maybe 100 points). So we won't expect big effect sizes. In comparing results from the same measure at different times, there is improvement over time.



*Sphericity.* SPSS says that Mauchly's test of sphericity has a p-value of  $p=.60$  for Measure, which means that we have not violated the sphericity assumption for this variable, but a p-value of  $p=.001$  for the interaction between Time and Measure, meaning the assumption of sphericity is violated. R says that sphericity is violated for the Measure (TestType) variable and for the two interactions between Time and Measure (TestType;  $p=.04$ ) and Group and Measure (TestType;  $p=.0009$ ). Use the Greenhouse-Geisser or Huynh-Feldt correction when reporting interactions and main effects.

*Interactions and main effects.* In SPSS, look at the "Test of Within-Subjects Effects" box, and in R, look at the lines under \$ANOVA directly under the `print(ColeLongModel1)` line. Results here are from SPSS but are nearly the same for R.

Time\*Measure\*Group:  $F_{2,56}=0.05$ ,  $p=.90$ , partial eta-squared=.002 (Greenhouse Geisser correction)

Time\*Measure:  $F_{2,56}=10.31$ ,  $p=.001$ , partial eta-squared=.27 (Greenhouse Geisser correction)

Measure\*Group:  $F_{2,56}=3.50$ ,  $p=.04$ , partial eta-squared=.11 (Greenhouse Geisser correction)

Time\*Group:  $F_{1,28}=0.24$ ,  $p=0.63$ , partial eta-squared=.008 (Greenhouse Geisser correction)

Measure:  $F_{2,56}=30.26$ ,  $p<.0005$ , partial eta-squared=.52 (Greenhouse Geisser correction)

Time:  $F_{1,28}=130.52$ ,  $p<.0005$ , partial eta-squared=.82 (Greenhouse Geisser correction)

Group:  $F_{1,28}=2.80$ ,  $p=.11$  (in SPSS this information is found later on in the output from the "Tests of Between-Subjects Effects" box).

Note: Remember that for the R output, the third part of the `ezANOVA()` output contains the epsilon value and p-value for the Greenhouse-Geisser and Huynh-Feldt corrections. In other words, GGe=Greenhouse-Geisser epsilon (which should be multiplied by the original error degree of freedom to get the correct df), p[GG] means the p-value of the Greenhouse-Geisser correction, HFe=Huynh-Feldt epsilon, and p[HF] is the p-value of the Huynh-Feldt correction.

The three-way interaction is not statistical and its effect size is essentially zero. The two-way interactions between Time and Measure and between Measure and Group are both statistical, with the partial eta-squared accounting for 27% of the variance for the first interaction and 11% in the second. With a statistical interaction, we don't really care about the main effects being statistical, so it doesn't much matter whether the three main effects for Measure, Time or Group are statistical. But our interactions tell us that the measures were different at different times (we saw this improvement over time on the measures in the descriptive stats) and that the groups did differently at different times (this would mean one group improved more over time than another one did), but the non-statistical result for the interaction between Time and Group means that the groups did not fare differently at the different times, which we already could see pretty clearly in the descriptive statistics because both groups improved over time, not just once.

In order to better understand this data we would need to do more analysis about how the factors interacted, but I will ask for further analysis of this data in Section 11.6.3 so I will stop here for now.

*Residual SSCP Matrix (SPSS only).* Look at the second row, the Covariance; here numbers along the diagonal should all be similar. This is not true (the lowest number, 40, is about half the highest number, 86), but they are not wildly different either. Off-diagonal elements should be equal. Again, this is not true but the numbers do not differ wildly. We will continue with the assumption that sphericity is violated.

*Assumptions (SPSS only).* The spread vs. level plots have only 2 points per graph so it is impossible to make any conclusions from them. In the residual plots examine the Std. Residual vs. Predicted table. For each variable the data are just two parallel lines, so again it's quite difficult to make any conclusions about whether the assumption of homogeneity of variances has been violated or not. Look at histograms of the 6 columns with RES\_ at the end of the dataset (GRAPHS > LEAGCY DIALOGS > HISTOGRAMS). These don't look too abnormal, so we might not worry too much that the data are not normally distributed.

For the Cook's distance numbers that were appended to the end of the dataset, we want to make sure none of these were over 1. The dataset is large so perhaps the fastest way to do this is to ask for descriptive stats for this: ANALYZE > DESCRIPTIVE STATISTICS > DESCRIPTIVES. Move over all the Cook's Distance variables. The mean score was only .04 with no scores over 1, which means there is no problem with outliers.

### 3 Larson-Hall (2004)

For SPSS use the LarsonHall2004.sav file. For R, use the lh2004Long dataset you created in Section 11.2.4.

### **SPSS Instructions:**

To conduct an RM ANOVA, use ANALYZE > GENERAL LINEAR MODEL > REPEATED MEASURES (remember, if you have previously split this file, you should go in and un-split it! The Data Editor has a little message at the bottom right that will say something like “Split by Cond” if your file is split. Go to DATA > SPLIT FILE and choose the “Analyze all cases” radio button). There is only one within-group factor, which is the different contrasts of the test (although here we will only test 3 different phonemic contrasts). Thus, go to the “Within-Subject Factor Name” box and label your variable CONTRAST and specify it has three levels in the “Number of Levels” box. ADD it, then click DEFINE. Move the 3 variables we are considering for this task (R\_L, SH\_SHCH, PJ\_P) into the “Within-subjects” box. Put the LEVEL variable in the “Between-subjects” box. Open the MODEL button and change to Type II SS. Create some interaction plots in the PLOTS button. Call for post-hocs on LEVEL (use LSD and Games-Howell) through the POSTHOC button. Open the SAVE button and tick “Cook’s Distance” and “Unstandardized.” Open the OPTIONS button and tick the boxes for “Descriptive statistics,” “Estimates of effect size,” “Residual SSCP matrix,” “Spread vs. Level plot,” and “Residual plot.” Click CONTINUE then OK.

### **R Instructions:**

```
library(ez)
```

```
lhModel1<-ezANOVA(data= lh2004Long, dv=.(score), wid=.(participant),  
within=.(contrast), between=.(level)) #main RM-ANOVA command
```

I get the following warning because groups are unequal, but I am OK with it.

Warning: Data is unbalanced (unequal N per group). Make sure you specified a well-considered value for the type argument to ezANOVA().

```
print(lhModel1)
```

```
(lhStats<-ezStats(data= lh2004Long, dv=.(score), wid=.(participant),  
within=.(contrast), between=.(level))) # descriptive statistics
```

## Results:

*Descriptive statistics.* First, we notice that the number of participants was modest, with a total  $N=41$  for four groups. Each sound has a possible 5 points, and I notice that overall, all of the proficiency levels did pretty well on the sounds, but the Beginners did seem to do a little worse, and the Native Russians (NR) do seem to have done the best. For R\_L, native Russians were better than all except the advanced group, while for SH\_SHCH, native Russians were only better than the Beginners. The PJ\_P contrast got the highest score in every group so there may be differences in how well people in different levels perform on the contrasts.

*Sphericity.* SPSS says that Mauchly's test of sphericity has a p-value of  $p=.11$  for Contrast, which means that we have not violated the sphericity assumption for this variable. R says that sphericity is not violated for either the Contrast variable ( $p=0.11$ ) or for the interaction between Level and Contrast ( $p=0.11$ ). Although it's not necessary according to the formal tests, the use of the Greenhouse-Geisser or Huynh-Feldt correction when reporting interactions and main effects might still be a good idea.

*Interactions and main effects.* In SPSS, look at the “Test of Within-Subjects Effects” box, and in R, look at the lines under \$ANOVA directly under the `print(lhLong2004)` line. Results here are from SPSS but are the same for R.

Contrast\*Level:  $F_{6, 74} = 2.78$ ,  $p = .017$ ,  $\text{partial-}\eta^2 = .18$

Contrast:  $F_{2, 74} = 21.48$ ,  $p < .0005$ ,  $\text{partial-}\eta^2 = .37$

Level:  $F_{3, 37} = 7.83$ ,  $p < .0005$ ,  $\text{partial-}\eta^2 = .39$  (in SPSS this information is found later on in the output from the “Tests of Between-Subjects Effects” box).

The two-way interaction between CONTRAST and LEVEL is statistical, meaning that not every group performed in parallel on all the contrasts, as might be expected given that native speakers were included. The main effects for CONTRAST and LEVEL are also statistical, but not important to us since we know the interaction is statistical.

In order to better understand this data we would need to do more analysis about how the factors interacted, but I will ask for further analysis of this data in Section 11.6.3 so I will stop here for now.

*Residual SSCP Matrix (SPSS only).* Look at the second row, the Covariance; here numbers along the diagonal should all be similar. This is not true (the lowest number, 0.14, is about one-fourth the highest number, 0.58), but they are not wildly different either. Off-diagonal elements should be equal. This holds fairly well. We will continue with the assumption that sphericity is upheld in this dataset.

*Assumptions (SPSS only).* The spread vs. level plots have only 4 points per graph so it is impossible to make any conclusions from them. In the residual plots examine the Std. Residual vs. Predicted table. The data do not look randomly scattered but neither do they have any other clear pattern, so it's quite difficult to make any conclusions about whether the assumption of homogeneity of variances has been violated or not. Look at histograms of the 3 columns with RES\_ at the end of the dataset (GRAPHS > LEGACY DIALOGS > HISTOGRAMS). These all look rather skewed, so we can say that data are not normally distributed.

For the Cook's distance numbers which were appended to the end of the dataset, we want to make sure none of these were over 1. The dataset is large so perhaps the fastest way to do this is to ask for descriptive stats for this: ANALYZE > DESCRIPTIVE STATISTICS > DESCRIPTIVES. Move over all the Cook's Distance variables. The mean score was only .02 with no scores over 1, which means there is no problem with outliers.

#### 4 Erdener and Burnham (2005)

For SPSS use the Erdener&Burnham2005.sav file. For R, use the **ErdenerLong** dataset you created in Section 11.2.4.

#### **SPSS Instructions:**

Choose ANALYZE > GENERAL LINEAR MODEL > REPEATED MEASURES. There are two within-group factors for this dataset—the condition the test was administered under (with 4 levels) and the target language (Spanish or Irish). Thus go to the “Within-Subject Factor Name” box and

label your first variable CONDITION and specify it has four levels in the “Number of Levels” box. ADD it. Go back to the “Within-Subject Factor Name” box and enter the next variable of TL, specify that it has two levels, click the ADD button, then click DEFINE. Move the 8 variables over, but notice that the target language should be 1 (Spanish) for the first entry, then 2 (Irish) for the second entry, so move the variables over in this order: SP\_AUDONLY, IRISH\_AUDONLY, SP\_AV, IRISH\_AV, SP\_AVORTH, IRISH\_AVORTH, SP\_AUDORTH, IRISH\_AUDORTH) into the “Within-subjects” box. Put the L1 variable in the “Between-subjects” box. Open the MODEL button and change to Type II SS. Create some interaction plots in the PLOTS button. There is no need to call for post-hocs on L1 because there are only two groups. Open the SAVE button and tick “Cook’s Distance” and “Unstandardized.” Open the OPTIONS button and tick the boxes for “Descriptive statistics,” “Estimates of effect size,” “Residual SSCP matrix” “Spread vs. Level plot,” and “Residual plot.” Click CONTINUE then OK.

### **R Instructions:**

```
library(ez)
```

```
ELModel1<-ezANOVA(data=ErdenerLong, dv=.(score), wid=.(participant),  
within=.(Condition, TL), between=.(l1)) #main RM-ANOVA command
```

When I run this I get this warning, and the model is not created:

```
Error in ezANOVA_main(data = data, dv = dv, wid = wid, within = within, :  
One or more cells returned NA when aggregated to a mean. Check your data.
```

My first suspicion is that I am missing data and have NAs in the file. The `ezAnova()` command cannot deal with any missing data. When I go back and look at the active data set in R



Commander, I see that I am not missing any data, but that some of the participants' scores are zero so it seems that there may be a problem with the model in that some means are simply zero.

I decide to check the descriptive statistics first:

```
(ErdenerStats<-ezStats(data=ErdenerLong, dv=.(score), wid=.(participant),  
within=.(Condition, TL), between=.(I1))) #descriptive statistics
```

This command returns the same warning (I get no results using it), so I go to R Commander and try STATISTICS > SUMMARIES > NUMERICAL SUMMARIES. The only variable is **score**, but it can be split by 3 variables in this dataset—**condition**, **I1**, and **TL**. This R Commander command will not split the variable by all three variables at one time, however. Suddenly, I think that maybe having the long form of the data is my problem, so I go back to the Erdener file as my active dataset and try the NUMERICAL SUMMARIES menu choice. Aha! This works, since my data are split up into their separate parts as separate columns here. I see that none of my means or SDs are zero, and there are no NAs in the data, and each division has N=31 or N=32 (nothing's too small) so I'm not sure what the problem is and don't know how I could fix it, but I decide to try to run the **ezANOVA()** command again with just one within-groups division:

```
ELModel1<-ezANOVA(data=ErdenerLong, dv=.(score), wid=.(participant),  
within=.(Condition), between=.(I1)) #main RM-ANOVA command
```

I get a new error message:

Warning: Collapsing data to cell means. \*IF\* the requested effects are a subset of the full design, you must use the “within\_full” argument, else results may be inaccurate.

I’m not sure what this means, so I look at the help files for this command:

?ezANOVA

It tells me that if I do have more than one within-group variable, I have to make sure to specify the entire structure of the dataset with another argument, `within_full`. So I try this.

```
ELModel1<-ezANOVA(data=ErdenerLong, dv=.(score), wid=.(participant),  
within=.(Condition), within_full=.(Condition, TL), between=.(I1)) #main RM-ANOVA
```

command

But no, this doesn’t work either. Originally, I was stumped as to what to do, but after doing Exercise #5, I realized that there probably were NAs in the data that I had missed. After all, it’s a very large file! This is the regular command you could use to check, but then you would have to scan through the very long list of “FALSE”s:

```
is.na(ErdenerLong)
```

So it’s more efficient to sum up the number of “TRUE”s like this (this is the code from this R Commander choice: STATISTICS > SUMMARIES > COUNT MISSING OBSERVATIONS):

```
sapply(ErdenerLong, function(x)(sum(is.na(x)))) # NA counts
```

Sure enough, I do find one NA! Since I do have a large number of participants, I decide to just remove that case. Here is how:

```
Erdener.compl<-ErdenerLong[complete.cases(ErdenerLong),]
```

The command says to make a new file where only the rows which are complete cases are found, so we are totally removing the row which has an NA. Will that be enough, or do I have to remove that one participant who has an NA entirely, as that person's data was repeated 8 times in the `ErdenerLong` dataset? I'll try to run the `ezANOVA()` command again with the new dataset:

```
ELModel1<-ezANOVA(data=Erdener.compl, dv=.(score), wid=.(participant),  
within=.(Condition, TL), between=.(I1))
```

It doesn't work. As I suspected, this is probably because the same participant whose row was removed still exists for the other variables. So I'll go back to the originally imported dataset and remove the NA from there first, then construct the long version of the file again.

```
library(reshape2) #open the library for the melt( ) command
```

```
erdener.compl<-erdener[complete.cases(erdener),] #get rid of the NA here
```

```
names(erdener.compl)
```

```
length(erdener.compl$I1)
```

```

erdener.compl<-cbind(erdener.compl, participant=factor(rep(c(1:63))))
names(erdener.compl) #verify the participant variable is now at the end of the dataset
ErdenerLong<-melt(erdener.compl,
id.vars=c("participant", "I1"),
value.name="score")
ErdenerLong<-cbind(ErdenerLong, colsplit(ErdenerLong$variable, "_", names=c("TL",
"Condition"))))
str(ErdenerLong)
ErdenerLong$variable <- NULL
ErdenerLong$TL<-as.factor(ErdenerLong$TL)
ErdenerLong$Condition<-as.factor(ErdenerLong$Condition)
write.table(ErdenerLong, "C:/Users/LarsonHall/Desktop/ErdenerLong.csv",
sep=";", col.names=TRUE, row.names=TRUE, quote=TRUE, na="NA")

```

I'm finished now and will run the NA query again:

```
sapply(ErdenerLong, function(x)(sum(is.na(x)))) # NA counts
```

I get zero NAs, so I'll try to run the `ezANOVA()` command again:

```

ELModel1<-ezANOVA(data=ErdenerLong, dv=.(score), wid=.(participant),
within=.(Condition, TL), between=.(I1)) #main RM-ANOVA command

```

The only error I get now is that the data is unbalanced, so I would go on to run the rest of the commands:

```
print(ELModel1)

(ELStats<-ezStats(data= ErdenerLong, dv=.(score), wid=.(participant),
within=.(Condition, TL), between=.(I1)))
```

If you have a problem like this and just can't seem to get the `ezANOVA()` command to run, you do have other options. You could use the R Commander menu choice of STATISTICS > MEANS > MULTI-WAY ANOVA. It will produce these commands in the R code:

```
AnovaModel.2 <- (lm(score ~ Condition*I1*TL, data=ErdenerLong))

Anova(AnovaModel.2)

with(ErdenerLong, (tapply(score, list(Condition, I1, TL), mean, na.rm=TRUE))) # means
with(ErdenerLong, (tapply(score, list(Condition, I1, TL), sd, na.rm=TRUE)))# std.
deviations

with(ErdenerLong, (tapply(score, list(Condition, I1, TL), function(x)
sum(!is.na(x))))) # counts
```

These commands will give me descriptive statistics and the ANOVA table of results but no test of sphericity and corrections if it is violated, as the `ezANOVA()` command does, and no tests for assumptions automatically included.

## Results:

*Descriptive statistics.* First, notice that there are 63/64 total participants in the study (the original number is 64 but participant 55 has a piece of missing data and SPSS deletes this entire case, giving 63 participants, while R works with what it has to give 64 participants). Notice from the descriptive stats that the English speakers get higher means than the Turkish speakers on every category except 2, which are Irish+Audio-visual+Orthography (Irish\_AVOrth) and Irish + Audio +Orthography (Irish\_AudOrth). The condition that seems to result in the highest scores for both languages is the Audio Only condition (except for the Turkish speakers in Irish, where their highest score is for Audio + Orthography).

*Sphericity.* SPSS says that Mauchly's test of sphericity has a p-value of  $p=.03$  for Condition, which means that we have violated the sphericity assumption for this variable. TL only has two levels so there is no need for sphericity. The interaction between Condition and TL has a p-value of .60, meaning the assumption of sphericity is not violated. Use the Greenhouse-Geisser or Huynh-Feldt correction when reporting interactions and main effects.

*Interactions and main effects.* In SPSS, look at the "Test of Within-Subjects Effects" box, in R using **ezANOVA** look under the \$ANOVA line, and in R Commander, look at the lines under **ANOVA(AnovaModel)**. Results here are from SPSS. R results with ezANOVA are the same as SPSS but results using the **Anova()** command are different in the particulars (different error term not given for each block, F value different, p-value different) but find the same results as to which terms are statistical except for the main effect of L1, which SPSS returns a p-value of  $p=.134$  for, while R returns a p-value of  $p=.0159$ .

Condition\*L1\*TL:  $F_{2,9,175.5}=4.18$ ,  $p=.008$ , partial eta-squared=.06 (Greenhouse Geisser correction)

Condition\* TL:  $F_{2,9,175.5}=11.68$ ,  $p<.0005$ , partial eta-squared=.16 (Greenhouse Geisser

L1\*TL:  $F_{1,61}=4.30$ ,  $p=.04$ , partial eta-squared=.07

TL:  $F_{1,61}=3.03$ ,  $p=.09$ , partial eta-squared=.05

Condition\*L1:  $F_{2,6,159.4}=1.82$ ,  $p=.15$ , partial eta-squared=.03 (Greenhouse Geisser)

Condition:  $F_{2,6,154.4}=24.02$ ,  $p<.0005$ , partial eta-squared=.28 (Greenhouse Geisser)

L1:  $F_{1,61}=2.30$ ,  $p=.13$ , partial eta-squared=.04 (in SPSS this information is found later on in the output from the “Tests of Between-Subjects Effects” box).

The three-way interaction is statistical, although its effect size is very small, accounting for only 6% of the variance in the model. However, because the three-way interaction is statistical we will mostly likely not be interested in any of the other effects by themselves, but instead want to understand how L1 interacts with TL and with the Condition. From the descriptives we could see that English L1 tends to result in higher scores for most categories, but not all, and also that some conditions result in higher mean scores than others, but this does not hold true for all L1s, so in order to better understand this data we would need to do more analysis about how the factors interact, but I will ask for further analysis of this data in Section 11.6.3 so I will stop here for now.

*Residual SSCP Matrix (SPSS only).* Look at the second row, the Covariance; here numbers along the diagonal should all be similar. This is not true (the lowest number, 2.4, is about one third the

highest number, 7.6), but they are not wildly different either. Off-diagonal elements should be equal. Again, this is not true but the numbers do not differ wildly. As we saw in the specific results, the variable of Condition violates sphericity, probably resulting in this matrix which shows some anomalies but is not as extreme as others.

*Assumptions (SPSS only).* The spread vs. level plots have only 2 points per graph so it is impossible to make any conclusions from them. In the residual plots examine the Std. Residual vs. Predicted table. For each variable the data are just two parallel lines, so again it's quite difficult to make any conclusions about whether the assumption of homogeneity of variances has been violated or not. Look at histograms of the 6 columns with RES\_ at the end of the dataset (GRAPHS > LEAGCY DIALOGS > HISTOGRAMS). Actually, since there are so many to look at and since you can only do one at a time through the SPSS menu, you might be interested here to use SPSS syntax. Put in your choices to do one histogram, but don't press OK. Instead press the "Paste" button. This will open up the SPSS Syntax Editor. You will see something like this code:

```
DATASET ACTIVATE DataSet10.
```

```
GRAPH
```

```
/HISTOGRAM=RES_1.
```

Just copy the whole thing 7 more times, then change the number "1" to numbers "2" through "8" to get histograms for the other 7 variables. Now choose RUN from the menu and pick ALL. You'll get all the histograms.



For most of these histograms the data are heavily skewed positively (more data to the left than would be expected by a normal distribution), so we conclude that the data are not normally distributed.

For the Cook's distance numbers which were appended to the end of the dataset, we want to make sure none of these were over 1. The dataset is large so perhaps the fastest way to do this is to ask for descriptive stats for this: ANALYZE > DESCRIPTIVE STATISTICS > DESCRIPTIVES. Move over all the Cook's Distance variables. The mean score was only .02 with no scores over 1, which means there is no problem with outliers.

## 5 Toth (2008)

For SPSS use the Toth2008Prod.sav file. For R, use the **tothProdLong** dataset you created in Section 11.2.4.

### **SPSS Instructions:**

To conduct an RM ANOVA, use ANALYZE > GENERAL LINEAR MODEL > REPEATED MEASURES. There is only one within-group factor, which is the different times of the test. Go to the "Within-Subject Factor Name" box and label your variable TIME and specify it has three levels in the "Number of Levels" box. ADD it, then click DEFINE. Move the 3 variables we are considering for this task (PRODPRETEST, PRODPOSTTEST, DELAYEDPOSTTEST) into the "Within-subjects" box. Put the GROUP variable in the "Between-subjects" box. Open the MODEL button and change to Type II SS. Create some interaction plots in the PLOTS button. Call for post-hocs on GROUP (use LSD and Games-Howell) through the POSTHOC button. Open the SAVE button and tick "Cook's

Distance” and “Unstandardized.” Open the OPTIONS button and tick the boxes for “Descriptive statistics,” “Estimates of effect size,” “Residual SSCP matrix,” “Spread vs. Level plot,” and “Residual plot.” Click CONTINUE then OK.

### **R Instructions:**

```
library(ez)
```

```
TothModel1<-ezANOVA(data= tothProdLong, dv=.(score), wid=.(subject),  
within=.(time), between=.(group)) #main RM-ANOVA command
```

I get the following warning and the model is not created:

```
Error in ezANOVA_main(data = data, dv = dv, wid = wid, within = within, :  
One or more cells returned NA when aggregated to a mean. Check your data.
```

This is the same warning I got in the previous exercise (#4). I’ll check the data through R Commander using the original **tothProd** dataset. I notice that there is one NA in the original dataset for one variable, so there is also one NA in the transformed long version of the dataset. Remember, you can look through the data, or you could also use this command to check for NAs:

```
sapply(tothProdLong, function(x)(sum(is.na(x)))) # NA counts
```

You'll find that there is an NA. This would probably be a great place to impute your data, since a person seems to be missing because they missed testing at one time period. Checking back to Section 1.5.3 that details how to impute data in R, I run the following commands:

```
library(mice)
```

```
imp<-mice(tothProdLong, 5)
```

```
impToth<-complete(imp)
```

```
names(impToth) #just to check that I have the right variables in this new dataset I've imputed
```

```
apply(impToth, function(x)(sum(is.na(x)))) #just to check there are no NAs now
```

Now will the `ezAnova( )` command run?

```
TothModel1<-ezANOVA(data= impToth, dv=.(score), wid=.(subject),
```

```
within=.(time), between=.(group)) #main RM-ANOVA command
```

Yes! I get warning about unbalanced data but the model is produced.

```
print>TothModel1)
```

```
(TothStats<-ezStats(data= impToth, dv=.(score), wid=.(subject),
```

```
within=.(time), between=.(group))) # descriptive statistics
```

## Results:

*Descriptive statistics.* First, we notice that the number of participants was a little larger than normal, with a total  $N=78$  for the three groups and about 25 participants in each group. I notice that the mean score increases quite a bit for both the Teacher-led and Learner-led group from pretest to posttest (and then slips back a little in the delayed posttest), but this is not true for the control group. As for comparing the groups among themselves at each time period, in the pretest they are all fairly similar but in the posttest and delayed posttest the Learner-led group and Teacher-led group seem quite similar.

*Sphericity.* SPSS says that Mauchly's test of sphericity has a p-value of  $p=.004$  for Time, which means that the sphericity assumption for this variable is violated. R says that sphericity is violated for both the Time variable ( $p=0.004$ ) and for the interaction between Time and Group ( $p=0.004$ ). Use the Greenhouse-Geisser or Huynh-Feldt correction when reporting interactions and main effects.

*Interactions and main effects.* In SPSS, look at the "Test of Within-Subjects Effects" box, and in R, look at the lines under \$ANOVA directly under the `print(TothModel1)` line. Results here are from SPSS but are similar for R (R used an imputed dataset while SPSS did not).

Time\*Group:  $F_{3.5, 130.01} = 15.20, p < .0005, \text{partial-}\eta^2 = .29$  (Greenhouse-Geisser)

Time:  $F_{1.8, 130.01} = 58.25, p < .0005, \text{partial-}\eta^2 = .44$  (Greenhouse-Geisser)

Group:  $F_{2, 74} = 50.38, p < .0005, \text{partial-}\eta^2 = .58$  (in SPSS this information is found later on in the output from the "Tests of Between-Subjects Effects" box).

The two-way interaction between TIME and GROUP is statistical, meaning that not every group performed in parallel on all the contrasts, as we expected from looking at the descriptives. The main effects for TIME and GROUP are also statistical, but not important to us since we know the interaction is statistical.

In order to better understand this data we would need to do more analysis about how the factors interacted, but I will ask for further analysis of this data in Section 11.6.3 so I will stop here for now.

*Residual SSCP Matrix (SPSS only).* Look at the second row, the Covariance; here numbers along the diagonal should all be similar. This is not true (the lowest number, 0.006, is about 12 times smaller than the highest number, 0.073). Off-diagonal elements should be equal and this is also not true. This confirms Mauchley's test that sphericity is violated.

*Assumptions (SPSS only).* The spread vs. level plots have only 3 points per graph so it is impossible to make any conclusions from them. In the residual plots examine the Std. Residual vs. Predicted table. Mostly the data look randomly scattered so I assume the assumption of homogeneity of variances has been upheld. Look at histograms of the 3 columns with RES\_ at the end of the dataset (GRAPHS > LEAGCY DIALOGS > HISTOGRAMS). The data for the pretest are heavily skewed and definitely not normally distributed, but for the posttest and delayed posttest residuals they may fall within the normal range of variation for normal distribution.

For the Cook's distance numbers which were appended to the end of the dataset, we want to make sure none of these were over 1. The dataset is large so perhaps the fastest way to do this is to ask for descriptive stats for this: ANALYZE > DESCRIPTIVE STATISTICS > DESCRIPTIVES. Move over all the Cook's Distance variables. The mean score was only .01 with no scores over 1, which means there is no problem with outliers.

### **11.6.3 Application Activities with Further Exploration of RM ANOVA Using Simple Interaction Effects and Simple Main Effects (SPSS and R Instructions)**

#### **1 Lyster (2004) Task Completion data**

For SPSS use the Lyster.Written.sav file. For R, use the **lyser.wide** dataset.

Look at the research questions:

- 1 Will FFI help students to assign French grammar more accurately (in other words, are all three groups that received FFI better than the Control group)?
- 2 Is FFI more effective with feedback (i.e. FFI recast and FFI prompt) than without (FFI only)?

Both of these questions will need to be answered by using planned comparisons. To use planned comparisons, we'll need to look at only one time period at a time. In other words, we basically need to use a one-way ANOVA to answer this question, and we'll only be able to answer it for one time period at a time. I will only look at the answers for the immediate posttest, and leave it to you to do it for the delayed posttest if you like. So the first step is to create gainscores for the

improvement from pretest to posttest.

### **a Calculate a Gainscore:**

#### **SPSS Instructions:**

TRANSFORM > COMPUTE VARIABLE. Create a “Numeric Expression” that says “POST1TASKCOMPL – PRETASKCOMPL”. In the box labeled “Target Variable,” give it a new name (GAINSCORE).

#### **R Instructions:**

```
lyster.wide$gainscore <- with(lyster.wide, post1taskcompl- pretaskcompl)
```

### **b Set up the Planned Comparisons Table and Run the Analysis**

There are 4 conditions. We have basically eliminated the time component by only examining the gain from the pretest to the posttest, so we just need to make sure which order the groups are in:

```
levels(lyster.wide$cond)
```

```
[1] "FFIrecast" "FFIprompt" "FFIonly" "Comparison"
```

So to compare all of the FFI groups to the control group (question #1), we’ll enter in the numbers “1, 1, 1, -3.” To compare the FFI recast and FFI prompt to the FFI only group, we’ll enter in the numbers “1, 1, -2, 0.”

#### **SPSS Instructions:**

With your Lyster.Written.sav dataset open, choose One-Way ANOVA (ANALYZE > COMPARE MEANS > ONE-WAY ANOVA) and put the GAINSCORE variable in the “Dependent List” box and COND in the “Factor” box. Open the CONTRASTS button and enter in the numbers for each of the

two questions as listed above. You'll enter the 4 numbers in order, putting each number by turn in the "Coefficients" box and pressing the "Add" button after each one. When you finish the first set of numbers, push the "Next" button to go on to the next comparison. Press Continue after entering the second set of numbers.

Open the Options button and tick the "Descriptive Statistics" box. Press Continue. Open the Bootstrap button and tick the "Perform Bootstrapping" box. Change the number of samples to 10,000 (or less, as this can get complicated and take a very long time to run). Change the "Confidence Intervals" to "BCa." Press Continue. Press OK to run your analysis for Questions #1 and #2.

In the output, you'll want to check the line with the contrasts (the code at the top) to make sure you filled in all of the coefficients correctly and that they match up with the contrasts you entered.

### **R Instructions:**

Create the planned comparison contrasts

```
contr<-rbind("Q1"=c(1,1,1,-3),  
"Q2"=c(1,1,-2,0))
```

To get the confidence intervals for these comparisons, first open the multcomp package and then create a regression model that models the gainscore as a function of the condition.



```
library(multcomp)
```

```
fit<-aov(gainscore~cond, data=lyster.wide)
```

Now call for the confidence intervals for the contrasts we specified above:

```
confint(glht(fit,linfct=mcp(cond=contr)))
```

```
plot(glht(fit,linfct=mcp(cond=contr))) #if you want, plot the contrasts too
```

```
summary(glht(fit,linfct=mcp(cond=contr))) #if you want p-values
```

Don't forget to get descriptive statistics too (here's one way):

```
with(lyster.wide, (tapply(gainscore, list(cond), mean, na.rm=TRUE))) # means
```

```
with(lyster.wide, (tapply(gainscore, list(cond), sd, na.rm=TRUE))) # sd
```

```
with(lyster.wide, (tapply(gainscore, list(cond), function(x) sum(!is.na(x))))) # counts
```

### c. Results

For Question #1, 95% CIs (from R) show that the difference in means between the three FFI groups and the control group is [10.42, 21.39], showing a very large difference between the FFI groups and the control group. From the descriptive statistics (which we have only for the individual groups, not for all three FFI groups versus the Control group) we see that the mean gainscores of the FFI groups are all much larger than the Control group's gainscores, so we assume that the answer to question of whether the FFI groups are better than the controls is yes, with the CI showing that we can be 95% sure that the difference is at least 10 points great, and maybe as much as 21 points greater, which shows a large effect size and some good precision. From SPSS we get the less informative data that this contrast is statistical,  $t(129.6)=7.65$ ,  $p<.0005$  (using the unequal variances line).

For Question #2, the results are not as strong, but still 95% CIs (from R) show that the difference in means between the two FFI groups with feedback and the FFI group without feedback is [5.44, 13.76], showing some difference between the two contrasts. The difference is not as large as the difference between the FFI groups and the control, but is clearly different from zero and, with 95% confidence, we assume at its worst to be at least 5 points of difference. The descriptive statistics show that the FFI Recast and FFI Prompt groups had higher mean scores than the FFI only group, so we assume that the statistical contrast shows that the FFI groups with feedback produced better results than the FFI only group. From SPSS we get the less informative data that this contrast is statistical,  $t(98.2)=5.19$ ,  $p<.0005$  (using the unequal variances line).

### Question #3

```
confint(glht(fit,linfct=mcp(cond="Tukey")))
```

```
plot(glht(fit,linfct=mcp(cond=contr))) #if you want, plot the contrasts too
```

```
summary(glht(fit,linfct=mcp(cond=contr))) #if you want p-values
```

For effect sizes, we don't easily have means and standard deviations for the new groupings we created, but we can use information about  $t$ -tests (the  $t$ -value and the  $df$ ) to calculate an effect size from an online calculator (SPSS gives both, although R only gives  $p$ -values). Using this method (and the  $t$  and  $df$  for unequal variances), for Question #1,  $d=1.34$ , and for Question #2,  $d=1.05$ , both quite large effect sizes.

### 3 Which type of feedback is more effective in FFI—recasts or prompts?

For this question we can just run a one-way ANOVA with normal contrasts, making sure not to adjust p-values, and pick out the contrast between FFI recasts and FFI prompts. I will show how to do that, although actually, it would have been fine to set up a third contrast with planned comparisons and have run that along with the other planned comparisons in the previous analysis. The coefficient numbers would be “1, -1, 0, 0.”

#### **SPSS Instructions:**

Again open the One-Way ANOVA menu choice (ANALYZE > COMPARE MEANS > ONE-WAY ANOVA). Your previous analysis will still be there, and the only thing you need to do is open the PostHoc button and tick the “LSD” box. Press OK to run the analysis. The information will be in the “Post Hoc Tests” box.

#### **R Instructions:**

Change the contrast from the planned contrasts to the “Tukey” contrast, which will compare all groups in the between-groups IV to all other groups. Here is the code:

```
confint(glht(fit,linfct=mcp(cond="Tukey")))
```

```
plot(glht(fit,linfct=mcp(cond="Tukey")))
```

```
summary(glht(fit,linfct=mcp(cond="Tukey")))
```

## Results:

This time we can get CIs from both SPSS and R, and can say that the 95% CI of the mean difference between the FFI Recast and FFI prompt group is [2.46, 7.94], meaning that there is a difference between the groups although it is not a large difference in points, and may be as small as 2 points, or only as large as 8 points.

For effect sizes, I can use an online calculator and enter the means and SDs for the two groups to get  $d=0.92$ , a large effect size, although smaller than the effect sizes of the previous comparisons.

## 2 Lyster (2004) Binary Choice data

For SPSS use the Lyster.Written.sav file. For R, use the `lyser.wide` dataset.

### a. Calculate a Gainscore

#### SPSS Instructions:

TRANSFORM > COMPUTE VARIABLE. Create a “Numeric Expression” that says “POST1BINARY – PREBINARY.” In the box labeled “Target Variable”, give it a new name (GAINSCOREB).

#### R Instructions:

```
lyser.wide$gainscoreB <- with(lyser.wide, post1binary- prebinary)
```

### **b. Set up the Planned Comparisons Table and Run the Analysis**

Basically, the planned comparisons are already set up and are exactly the same as in Exercise #1.

This time, let's just add the one more planned comparison for research question #3 to the first two research question planned comparison.

Question 1: "1, 1, 1, -3"

Question 2: "1, 1, -2, 0"

Question 3: "1, -1, 0, 0"

#### **SPSS Instructions:**

With your Lyster.Written.sav dataset open, choose One-Way ANOVA (ANALYZE > COMPARE MEANS > ONE-WAY ANOVA).

If you've just done Exercise #1, you won't need to replicate everything below, just change the variable in the "Dependent List" box to the GAINSCOREB variable and open up the Contrasts button, hit "Next," and enter the third set of coefficients.

If this is the first time you are opening up the One-Way ANOVA command, put the GAINSCOREB variable in the "Dependent List" box and COND in the "Factor" box. Open the CONTRASTS button and enter in the numbers for each of the two questions as listed above. You'll enter the 4 numbers in order, putting each number by turn in the "Coefficients" box and pressing the "Add" button after each one. When you finish the first set of numbers, push the "Next" button to go on to the next comparison. Continue until all three sets of numbers are entered and then press CONTINUE.

Open the Options button and tick the “Descriptive Statistics” box. Press CONTINUE. Open the Bootstrap button and tick the “Perform Bootstrapping” box. Change the number of samples to 10,000 (or less, as this can get complicated and take a very long time to run; I used 2000 here). Change the “Confidence Intervals” to “BCa.” Press CONTINUE. Press OK to run your analysis.

In the output, you’ll want to check the line with the contrasts (the code at the top) to make sure you filled in all of the coefficients correctly and that they match up with what you wanted to enter.

### **R Instructions:**

Create the planned comparison contrasts

```
contr<-rbind("Q1"=c(1,1,1,-3),  
"Q2"=c(1,1,-2,0), "Q3"=c(1, -1, 0, 0))
```

To get the confidence intervals for these comparisons, first open the multcomp package and then create a regression model that models the gainscore as a function of the condition.

```
library(multcomp)  
fit<-aov(gainscoreB~cond, data=lyster.wide)
```

Now call for the confidence intervals for the contrasts we specified above:

```
confint(glht(fit,linfct=mcp(cond=contr)))
```

```
plot(glht(fit, linfct=mcp(cond=contr))) #if you want, plot the contrasts too
```

```
summary(glht(fit, linfct=mcp(cond=contr))) #if you want p-values
```

Don't forget to get descriptive statistics too (here's one way):

```
with(lyster.wide, (tapply(gainscoreB, list(cond), mean, na.rm=TRUE))) # means
```

```
with(lyster.wide, (tapply(gainscoreB, list(cond), sd, na.rm=TRUE))) # sd
```

```
with(lyster.wide, (tapply(gainscoreB, list(cond), function(x) sum(!is.na(x))))) # counts
```

### c. Results

For Question #1, 95% CIs (from R) show that the difference in means between the three FFI groups and the control group is [5.49, 20.35], showing some difference between the FFI groups and the control group, but a wider and less precise CI than was found for Exercise #1, and one that does not show as large as an effect since the lower level of the CI is closer to zero than it was for Question #1. From the descriptive statistics (which we have only for the individual groups, not for all three FFI groups versus the Control group) we see that the mean gainscores of the FFI groups are all larger than the Control group's gainscores (although not by much for FFI only), so we assume that the answer to question of whether the FFI groups are better than the controls is yes, with the CI showing that we can be 95% sure that the difference is at least 5 points great, and maybe as much as 20 points greater. From SPSS we get the less informative data that this contrast is statistical,  $t(95.5)=4.36$ ,  $p<.0005$  (using the unequal variances line). SPSS does not return bootstrapping values for this contrast. For effect sizes, I can use an online calculator and enter the  $t$ -value and df to get  $d=0.89$ .

For Question #2, the results are similar. 95% CIs (from R) show that the difference in means between the two FFI groups with feedback and the FFI group without feedback is [6.59, 17.86],

showing some difference between the two contrasts. The difference is clearly different from zero and, with 95% confidence, we assume at its worst to be at least about 6 points of difference. The descriptive statistics show that the FFI Recast and FFI Prompt groups had higher mean scores than the FFI only group, so we assume that the statistical contrast shows that the FFI groups with feedback produced better results than the FFI only group. From SPSS we get the less informative data that this contrast is statistical,  $t(84.8)=5.36$ ,  $p<.0005$  (using the unequal variances line). For effect sizes, I can use an online calculator and enter the  $t$ -value and  $df$  to get  $d=1.16$ .

For Question #3 we find a 95% CI very close to zero:  $[-7.42, -0.95]$ . The difference between the FFI Recast and FFI Prompt group could be quite small, just about 1 point of difference, with the descriptive statistics showing that the group with the better performance is the FFI Prompt group. The CI is not too wide, and shows us that the upper limit of the CI is only about 7 points, which is meaningful but not a huge difference. For SPSS we can use the bootstrapped CI under the “Post Hoc Tests” to say that the 95% CI is  $[-6.83, -1.54]$ , very similar to the results from R. We can also say, with less information but giving us a  $t$ -value and a  $df$  to use for the effect size, that the difference is statistical,  $t(84.5)=-3.11$ ,  $p=.003$  (using the unequal variances line). For effect sizes, I use an online calculator and enter the means and SDs to get  $d=0.66$ .

### 3 Cole (1927)

For SPSS use the Cole1927.sav file. For R, use the **ColeLong** dataset you created in Section 11.2.4.



Remember that a simple interaction effect is looking at the effects of the other variables at one level of a third variable. If we split up the variable of language measure, which has 3 levels, then we will be looking at the effects of group, with two levels, and testing time, with two levels. If we are interested in whether the groups differed at Time 1 and Time 2, we could proceed as outlined below. If we are simply interested in which group gained more from Time 1 to Time 2, it would be easiest to just calculate a gainscore and, since there are only two groups, run an independent-samples t-test for the two groups (in this way we'd be able to get confidence intervals for the mean difference between groups as well).

### **SPSS Instructions:**

ANALYZE > GENERAL LINEAR MODEL > REPEATED MEASURES. Go to the "Within-Subject Factor Name" box and add the variable of TIME and specify it has two levels in the "Number of Levels" box. ADD it, then click DEFINE. Move the 2 variables (TIME1\_VOCAB, TIME2\_VOCAB) into the "Within-subjects" box. Put the GROUP variable in the "Between-subjects" box. Open the MODEL button and change to Type II SS. Press Continue. Open the PLOTS button and ask for one plot with GROUP and TIME. Press ADD, then CONTINUE. No post-hocs are needed as there are only two groups. Open the SAVE button and tick "Cook's Distance" and "Unstandardized." Open the OPTIONS button and tick the boxes for "Descriptive statistics," "Estimates of effect size," "Residual SSCP matrix," "Spread vs. Level plot," and "Residual plot." Click CONTINUE. You could run the analysis, and if the interaction between the two variables is statistical, you could get further information about this interaction by doing the following:

Open the OPTIONS button and move the Group\*time interaction to the right, to the “Display Means For” box. Click CONTINUE. Open the Paste button. Because you asked for the means of the GROUP\*TIME interaction, in the syntax you’ll see a line like this:

```
/EMMEANS=TABLES(Group*time)
```

Copy this line twice, and add an additional part at the end to compare GROUPS and compare TIME, like this (paste these in under the above line):

```
/EMMEANS=TABLES(Group*time)COMPARE(Group)
```

```
/EMMEANS=TABLES(Group*time)COMPARE(time)
```

Choose RUN > ALL.

Now you’ll see a box that says “Pairwise Comparisons” where you can break down the interaction between Time and Group.

Repeat for the other gainscores.

### **R Instructions:**

```
library(ez)
```

To use the RM ANOVA we need the long form of the data, but to test only one language measure at a time we’ll need to split the data first before using the `ezANOVA()` command.

```
ColeVocab <- subset(ColeLong, subset=TestType=="vocab_percent")
```

```
ColeGrammar <- subset(ColeLong, subset=TestType=="grammar_percent")
```

```
ColeReading <- subset(ColeLong, subset=TestType=="reading_percent")
```

Now run an RM Anova on just the Vocab data:

```
ColeVModel1<-ezANOVA(data=ColeVocab, dv=.(score), wid=.(participant),
```

```
within=.(time), between=.(group)) #main RM-ANOVA command
```

```
print(ColeVModel1)
```

```
(vocabStats<-ezStats(data= ColeVocab, dv=.(score), wid=.(participant),
```

```
within=.(time), between=.(group))) # descriptive statistics
```

If the time\*group interaction is statistical, I'll want some more testing between the variables, so I could use this command:

```
pairwise.t.test(ColeVocab$score, ColeVocab$time:ColeVocab$group,
```

```
p.adjust.method="fdr")
```

However, this would only return p-values for these comparisons.

For this dataset, I'm interested in confidence intervals of the difference in scores between Time 1 and Time 2 for each group (in other words, what is the interval where I can be 95% confident that their true gainscore lies in?). If there were more than 2 levels of the variable I could use a one-way ANOVA, but since there are only two levels, I have to run an independent-samples *t*-test. I'm interested in looking at how Time affected scores for each group, so I'll divide my data

further into each of the two groups, and use a  $t$ -test to look at the difference between Time 1 and Time 2, and that way I can get the confidence interval for the difference between the two scores.

I do it like this:

```
levels(ColeVocab$group)
```

```
[1] "Composition" "Grammar Translation"
```

```
ColeVocabG1<-subset(ColeLong, subset=group=="Composition")
```

```
ColeVocabG2<-subset(ColeLong, subset=group=="Grammar Translation")
```

Now run a  $t$ -test (in R Commander, Statistics > Means > Independent Samples T-Test, choose **time** as "Group" and **score** as "Response variable"):

```
t.test(score~time, alternative='two.sided', conf.level=.95, var.equal=FALSE,  
data=ColeVocabG1)
```

### Results:

*Vocab*: Descriptives show that at both time periods, the Composition group's scores were about 5 points higher than the Grammar Translation group, and both sets of scores went up in parallel at Time 2 (a means plot makes this quite clear). The interaction between time and group is not statistical and has basically no effect size ( $F_{1,28}=.041$ ,  $p=.84$ , partial eta-squared=.001), nor is the main effect of group statistical, although it has some small effect ( $F_{1,28}=3.25$ ,  $p=.082$ , partial eta-squared=.10). However, the main effect of Time is statistical ( $F_{1,28}=145.3$ ,  $p<.0005$ , partial eta-squared=.84), and the descriptives show that Time 2 had higher scores than Time 1 (although I

called for data to look at assumptions for this test in SPSS, I won't report on the assumptions here).

In SPSS 95% CIs (found in the Pairwise Comparisons box gotten by using the SPSS syntax) show that the difference in scores from Time 1 to Time 2 for the composition group was [9.87, 16.26] and for the grammar translation group it was [10.32, 16.70]. These are very similar CIs for the improvement over time, and show that both groups, with 95% confidence, improved at least 10% from Time 1 to Time 2, but could have improved as much as 16% (remember, scores are calculated in percents in the original file).

*Grammar:* Descriptives show that at both time periods and for both groups, scores do not differ much. There is some small improvement (about 5%) from Time 1 to Time 2 for both groups again (means plots show the parallelism clearly). The interaction between time and group is not statistical and has no effect size ( $F_{1,28}=.002$ ,  $p=.97$ , partial eta-squared=.000), nor is the main effect of group statistical ( $F_{1,28}=.027$ ,  $p=.61$ , partial eta-squared=.01). However, the main effect of Time is statistical ( $F_{1,28}=13.54$ ,  $p=.001$ , partial eta-squared=.33), and descriptives show Time 2 had higher scores than Time 1.

In SPSS 95% CIs (found in the Pairwise Comparisons box gotten by using the SPSS syntax) show that the difference in scores from Time 1 to Time 2 for the composition group was [1.14, 10.06] and for the grammar translation group it was [1.27, 10.20]. These are very similar CIs for the improvement over time, and show that both groups, with 95% confidence, improved only about at least 1% from Time 1 to Time 2, but could have improved as much as 10% (remember,

scores are calculated in percents in the original file). The lower limit shows that the improvement may be quite small.

*Reading:* Descriptives show that the composition group scored higher at Time 1 than the grammar translation group, and that both groups gained about 5% at Time 2 but both groups improved. The interaction between time and group is not statistical and has basically no effect size ( $F_{1,28}=.23$ ,  $p=.64$ , partial eta-squared=.008), but the main effect of group is statistical ( $F_{1,28}=4.74$ ,  $p=.04$ , partial eta-squared=.15) and descriptives show the composition group scored more highly than the grammar translation group. In addition, the main effect of Time is statistical ( $F_{1,28}=9.19$ ,  $p=.005$ , partial eta-squared=.25), and descriptive show Time 2 had higher scores than Time 1.

In SPSS 95% CIs (found in the Pairwise Comparisons box gotten by using the SPSS syntax) show that at times the difference in scores from Time 1 to Time 2 for the composition group was  $[-.51, 8.13]$ , meaning that there may be no statistical improvement for the composition group from Time 1 to Time 2 as the CI passes through zero. However, it is close to zero and might be narrowed and have greater precision with more participants, but it looks like the effect is not very strong. For the grammar translation group the 95% CI in change from Time 1 to Time 2 was  $[-.92, 9.56]$ , meaning that although there is a statistical difference between groups, the CI is very close to 0 in its lower limit and does not show large effects.

#### 4 Larson-Hall (2004)

For SPSS use the LarsonHall2004.sav file. For R, use the **lh2004Long**

dataset you created in Section 11.2.4.

In the RM ANOVA, the interaction between Contrast and Level was statistical, as were the main effects of Contrast and Level. To explore which contrasts and which language proficiency levels were different from one another, we'll break down the data for each of the three contrasts and for each of the four levels.

### **SPSS Instructions:**

Follow the instructions for running an RM ANOVA given in Exercise #3 for Section 11.5.4. This time, open up the Options button and move the interaction, LEVEL\*PHONEMIC, to the right, to the box labeled "Display Means for:." Press CONTINUE. In the main dialogue box, press the Paste button and open the SPSS syntax. Add the following lines in red below the first line:

```
/EMMEANS=TABLES(level*Phonemic)
```

```
/EMMEANS=TABLES(level*Phonemic)COMPARE(level)
```

```
/EMMEANS=TABLES(level*Phonemic)COMPARE(Phonemic)
```

Go to the menu and choose RUN > ALL. In the output, under "2.level\*Phonemic" you will see a "Pairwise Comparisons" box. Further down, in "3. level\*Phonemic" you will see another "Pairwise Comparisons" box. This is the place I got my 95% CIs from.

### **R Instructions:**

To look at how the different proficiency levels performed on the 3 contrasts, divide up the **lh2004Long** file into the 4 proficiency levels.

```
LHBeg <- subset(lh2004Long, subset=level=="Beginner")
```

```
LHInt <- subset(lh2004Long, subset=level=="Intermediate")
```

```
LHAdv <- subset(lh2004Long, subset=level=="Advanced")
```

```
LHNR <- subset(lh2004Long, subset=level=="Native Russian")
```

Now run one-way ANOVAs on each of these files. If using R Commander (don't forget to change the active dataset), go to STATISTICS > MEANS > ONE-WAY ANOVA. Choose **contrast** for the "Groups" and **score** for the "Response Variable." Tick the "pairwise comparison of means" box. Press OK. Here is the R code for this menu choice:

```
AnovaModel.4 <- aov(score ~ contrast, data=LHBeg)
```

```
summary(AnovaModel.4)
```

```
with(LHBeg, numSummary(score, groups=contrast, statistics=c("mean", "sd")))
```

```
local({
```

```
.Pairs <- glht(AnovaModel.4, linfct = mcp(contrast = "Tukey"))
```

```
print(summary(.Pairs)) # pairwise tests
```

```
print(confint(.Pairs)) # confidence intervals
```

```
print(cld(.Pairs)) # compact letter display
```

```
old.oma <- par(oma=c(0,5,0,0))
```

```
plot(confint(.Pairs))
```

```
par(old.oma)
```

```
})
```



Repeat with each of the files to get CIs something like those in the Results section. To test differences between language proficiency levels for each contrast, divide up the data by contrast and then run one-way ANOVAs with level as the independent variable (“Groups” box in R Commander dialogue).

## Results:

I have created tables with the 95% CIs from SPSS in them:

Level	This vs.	This	95% CI
Beginner	R_L	PJ_P	-1.62, -0.75
	R_L	SH_SHCH	-0.56, -.56
	SH_SHCH	PJ_P	.74, 1.62
Intermediate	R_L	PJ_P	-1.42, -0.58
	R_L	SH_SHCH	-1.12, -0.05
	SH_SHCH	PJ_P	-0.003, 0.84
Advanced	R_L	PJ_P	-1.06, -0.14
	R_L	SH_SHCH	-0.68, 0.48
	SH_SHCH	PJ_P	0.04, 0.96
Native Russian	R_L	PJ_P	-0.64, 0.39
	R_L	SH_SHCH	-0.78, 0.53
	SH_SHCH	PJ_P	-0.51, 0.51

For the native Russians, all of the CIs pass through zero, so none of the contrasts are any harder than any of the others. For the learners, however, there are some contrasts that are more difficult than others. Looking at the CIs and the mean scores, I deduce that for the Advanced learners, R\_L and SH\_SHCH were harder than PJ\_P, although the CIs show that the lower limit of the difference was very small, so the effect was not large. For the Intermediate learners, strangely enough because mean scores are quite different, R\_L was not more difficult than PJ\_P, but this is probably because the SD of the R\_L contrast was quite a bit larger than the PJ\_P SD, and so this would be a good candidate for a robust examination. For the Intermediates, R\_L was more

difficult than SH\_SHCH, although again the lower level of the CI shows there might be very little difference. Lastly, for the Beginners, R\_L and SH\_SHCH were harder than PJ\_P by at least .75 of a point (with only 5 points for each category that is a good-sized difference). One would also be able to calculate effect sizes for each comparison by using an online calculator and entering the mean and standard deviation for each contrast (Beginner R\_L and Beginner PJ\_P, for instances, for the first row of the above table), but I haven't reported on that here.

The SPSS output also gives 95% CIs for differences between language proficiency levels for each contrast, and I have outlined how to do this for R. I will not give that answer here but trust that if the reader has followed the directions they will have this information as well.

## 5 Erdener and Burnham (2005)

For SPSS use the Erdener&Burnham2005.sav file. For R, use the **ErdenerLong** dataset you created in Section 11.2.4.

### 1 Do all of the participants perform better in conditions with visual information?

Remember there are 4 conditions: 1) Audio only; 2) Audio + Visual; 3) Audio + Visual + Orthography; 4) Audio + Orthography. Check on the order of the conditions in your file.

In the wide form of the data (Erdener&Burnham2005.sav file or previously imported into R as **erdener**) we essentially want to

#### **SPSS Instructions:**

The wide form of the data in the Erdener&Burnham2005.sav file is not going to be right to do this comparison. Basically, we have to create the long form of the data, which we practiced doing in R, but not in SPSS. In SPSS I recommended just doing this manually. So . . . manually put each of the two columns which contain the same condition (such as SP\_AUDONLY and IRISH\_AUDONLY) together. We won't bother to add an index column because in this analysis we're not splitting those up. Give a new label to each column and save the file with a different name. We're not done yet, though! Now stack all 4 columns together (so put AV under AudioOnly, etc.; a good idea is to put the correct condition number at the beginning when you copy in each new column so you know where your index values should change) and call this variable SCORE, then create an index column labeled CONDITION that labels each of the conditions as 1, 2, 3, or 4; I used my old L1 column, and cut and pasted the "1"s and "2"s that were already there. Go into the "Variable View" tab, "Values" column, and label those numbers with the correct condition. I then saved the file as ErdenerLong.sav. In my new file, the conditions are in the same order as listed under the first question, so to do a planned comparison I will use the contrast coefficients like this: -1, 1, 1, -1.

Choose One-Way ANOVA (ANALYZE > COMPARE MEANS > ONE-WAY ANOVA) and put the SCORE variable in the "Dependent List" box and CONTRAST in the "Factor" box. Open the CONTRASTS button and enter in the contrast coefficients for the question as listed above. You'll enter the 4 numbers in order, putting each number by turn in the "Coefficients" box and pressing the "Add" button after each one. When you finish press Continue.

Open the Options button and tick the “Descriptive Statistics” box. Press CONTINUE. Press OK to run your analysis.

In the output, you’ll want to check the Contrast Coefficients box to make sure you filled in all of the coefficients correctly and that they match up with the contrasts you entered.

### **R Instructions:**

Make sure what order the conditions are in:

```
levels(ErdenerLong$Condition)
```

```
[1] "audonly" "audorth" "av" "avorth"
```

Create the planned comparison contrasts:

```
contr<-rbind("Q1"=c(-1, -1, 1, 1))
```

To get the confidence intervals for these comparisons, first open the **multcomp** package and then create a regression model that models the **score** as a function of the **Condition**.

```
library(multcomp)
```

```
fit<-aov(score~Condition, data=ErdenerLong)
```

Now call for the confidence intervals for the contrasts we specified above:

```
confint(glht(fit,linfct=mcp(Condition=contr)))
```

```
summary(glht(fit,linfct=mcp(Condition=contr))) #if you want p-values
```

Don't forget to get descriptive statistics too (here's one way, but it will only give the scores for all 4 conditions, not the combined conditions):

```
with(ErdenerLong, (tapply(score, list(Condition), mean, na.rm=TRUE))) # means
```

```
with(ErdenerLong, (tapply(score, list(Condition), sd, na.rm=TRUE))) # sd
```

```
with(ErdenerLong, (tapply(score, list(Condition), function(x) sum(!is.na(x))))) # counts
```

### **Results:**

There is a statistical difference between the groups who saw visual cues and those who didn't,  $t(383.4) = -2.82$ ,  $p = .005$  (not assuming equal variances from SPSS). The 95% CI (from R) is [-2.00, -0.56], which is quite a way away from zero and given the size of the mean scores (around 2 or 3 points) is a large difference. Looking at mean scores for the 4 groups, the groups with visual parts in them are lower scores, which means fewer errors were made, so conclude that visual data made the production of nonwords more accurate. To calculate effect sizes I would have to split up groups differently, so I am not going to include that here!

## **2 How does L1 interact with target language on scores?**

We know from the results in Section 11.5.4 that the interaction between L1 and TL is statistical. But how do they interact? In SPSS use an RM ANOVA with syntax for comparing groups, but in R use an independent samples t-test.

### **SPSS Instructions:**

Go back to the Erdener&Burnham2005.sav file. Follow the instructions for running an RM ANOVA given in Exercise #4 for Section 11.5.4, but don't open any buttons. This time, open up the Options button and move the interaction, L1\*TL, to the right, to the box labeled "Display Means for:." While you're there, make sure to tick the "Descriptive statistics" box. Press CONTINUE. In the main dialogue box, press the Paste button and open the SPSS syntax. Add the following lines in red below the first line:

```
/EMMEANS=TABLES(L1*TL)
```

```
/EMMEANS=TABLES(L1*TL)COMPARE(L1)
```

```
/EMMEANS=TABLES(L1*TL)COMPARE(TL)
```

Go to the menu and choose RUN > ALL. In the output you will see "Pairwise Comparisons" boxes. These are the places I got my 95% CIs from.

### **R Instructions:**

To look at how the different L1s performed on target languages, divide up the **ErdenerLong** file into the 2 L1s.

```
Turkish <- subset(ErdenerLong, subset=l1=="Turkish")
```

```
English <- subset(ErdenerLong, subset=l1=="English")
```

Now run a *t*-test (in R Commander, STATISTICS > MEANS > INDEPENDENT SAMPLES T-TEST, choose **TL** as "Group" and **score** as "Response variable"):

```
t.test(score~TL, alternative='two.sided', conf.level=.95, var.equal=FALSE,  
data=Turkish)
```

Repeat for the other dataset, **English**. To examine the question from the other way around (whether there is a difference between Turkish and English L1 speakers on the target language of English or Spanish), divide the data up by TL and run t-tests.

### **Results:**

(from R) For the Turkish L1 speakers, they perform better if the TL is Irish than Spanish, 95% CI [0.14, 1.19]. For the English L1 speakers, there is no difference if the TL is Spanish or Irish, 95% CI [-0.60, 0.47]. (from SPSS) For the TL of Spanish, there is a difference in how well English L1 and Turkish L1 speakers perform, with English speakers performing better, 95% CI [.14, 1.50]. For the TL of Turkish, there is no difference in how English L1 and Turkish L1 speakers perform, 95% CI [-.62, 0.79]. Using an online effect size calculator, the comparison between

## **6 Toth (2008)**

For SPSS use the Toth2008Prod.sav file. For R, use the **tothProdLong** dataset you created in Section 11.2.4.

### **SPSS Instructions:**

With Toth2008Prod.sav dataset open, choose One-Way ANOVA (ANALYZE > COMPARE MEANS > ONE-WAY ANOVA) and put the PRODPOSTTEST variable in the “Dependent List” box and GROUP in the “Factor” box. Open the PostHoc button and tick LSD. Press CONTINUE. Open the

Options button and tick the “Descriptive Statistics” box. Press CONTINUE. Press OK. You’ll look for the pairwise comparisons in the box called “Multiple Comparisons” under the Post Hoc Tests heading.

### **R Instructions:**

In R Commander, open STATISTICS > MEANS > ONE-WAY ANALYSIS OF VARIANCE. Choose **group** for the “Groups” box and **prodposttest** for the “Response Variable.” Tick the “Pairwise comparison of means” box. Press OK. The R code for this is:

```
AnovaModel.7 <- aov(prodposttest ~ group, data=tothProd)
summary(AnovaModel.7)
with(tothProd, numSummary(prodposttest, groups=group, statistics=c("mean",
"sd")))
local({
.Pairs <- glht(AnovaModel.7, linfct = mcp(group = "Tukey"))
print(summary(.Pairs)) # pairwise tests
print(confint(.Pairs)) # confidence intervals
print(cld(.Pairs)) # compact letter display
old.oma <- par(oma=c(0,5,0,0))
plot(confint(.Pairs))
```



```
par(old.oma)
```

```
))
```

### Results:

Non-adjusted pairwise comparisons from SPSS show that all the groups differ from each other, with mean scores showing that the Teacher-led group was the best, Learner-led was second, and Control was last. Here are the 95% CIs for the difference between groups, along with effect sizes calculated using an online calculator.

Comparison	95% CI	Effect size
Teacher led_ Learner-led	.01, .30	.47
Teacher led_ Control	.49, .78	2.78
Learner led_ Control	.33, .63	2.12

It's clear that having some focus on the grammatical form of Spanish clitics was highly useful to the groups, and resulted in large effect sizes. There was a difference between the two experimental groups, with the Teacher-led group outperforming the Learner-led group, but the effect size wasn't as large as just getting some instruction.