Appendix E: R Script to Reproduce Analysis

E1. Loading the CSV files into R

```
#Packages Used (in the end some of these are probably not necessary
but we are not sure which ones)
library(dplyr)
library(tidyr)
library(ggplot2)
library(psych)
library(splitstackshape)
library(car)
library(stargazer)
library(hexbin)
library(Hmisc)
library(arm)
library(effsize)

#Set your working directory to location of CSV files and all R-scripts
setwd("……")

#We load all the scripts from a folder named Rscripts, create this
folder in your working directory or change the file name saves

#Load CSV files into R
PreIntervention <- read.csv("Intervention_Presurvey.csv")
PostIntervention <- read.csv("Intervention_Postsurvey.csv")
FollowIntervention <- read.csv("Intervention_Followupsurvey.csv")
NonIntervention <- read.csv("Nonintervention_Presurvey.csv")
```

E2. Creating the necessary dataframes

```
#Record Questions in Dataframe
Questions <- data.frame(paste("Q",1:22,
sep=""),names(PreIntervention[,8:29]))
colnames(Questions) <- c("ID","Question Text")

#Change column names for questions to Q1-Q22
#We are doing this because the full questions are unwiedingly long
colnames(PreIntervention)[8:29] <- paste("Q", 1:22, sep="")
colnames(PostIntervention)[8:29] <- paste("Q", 1:22, sep="")
colnames(FollowIntervention)[8:29] <- paste("Q", 1:22, sep="")
colnames(NonIntervention)[8:29] <- paste("Q", 1:22, sep="")

#Simplifying and Putting into English Demographic Column Names
colnames(PreIntervention)[3:6] <- c("Age","Origin","Degree","Year")
colnames(PostIntervention)[3:6] <- c("Age","Origin","Degree","Year")
colnames(FollowIntervention)[3:6] <- c("Age","Origin","Degree","Year")
colnames(NonIntervention)[3:6] <- c("Age","Origin","Degree","Year")

#Creating a vector of names we can use for graphs
```

```
xaxis_names <-c(paste("Q",1:22, sep=""),"D","P","S","B")

#Next we are going to create composite variables for each "Knowledge
Domain"
#These vectors identify which questions go in each domain
DeclarativeVec <- c("Q1","Q2","Q3","Q4","Q5","Q16","Q17")
ProceduralVec <- c("Q6","Q10","Q11","Q12")
SubjectiveVec <- c("Q7","Q8","Q9","Q13","Q14","Q15")
BehaviorVec <- c("Q18","Q19","Q20","Q21","Q22")

#Print Table
Indexes <- rbind(paste(DeclarativeVec,collapse=",
"),paste(ProceduralVec,collapse=", "),paste(SubjectiveVec,collapse=",
"),paste(BehaviorVec,collapse=", "))
row.names(Indexes) <-
c("Declarative","Procedural","Subjective","Behavior")
stargazer(Indexes,summary=FALSE ,title="Questions included in each
Index for Analysis", colnames=FALSE, out =
"Rresults/indexquestions.html")

#Now we add a calculated column to each dataset with these composite
variables
#Calculated by the mean of the responses for the identified questions
#We omit any empty responses
PreIntervention <-
transform(PreIntervention,Declarative=round(rowMeans(PreIntervention[D
eclarativeVec],na.rm=TRUE),1),

Procedural=round(rowMeans(PreIntervention[ProceduralVec],na.rm=TRUE),1
),

Subjective=round(rowMeans(PreIntervention[SubjectiveVec],na.rm=TRUE),1
),

Behavior=round(rowMeans(PreIntervention[BehaviorVec],na.rm=TRUE),1))
PostIntervention <-
transform(PostIntervention,Declarative=round(rowMeans(PostIntervention
[DeclarativeVec],na.rm=TRUE),1),

Procedural=round(rowMeans(PostIntervention[ProceduralVec],na.rm=TRUE),
1),

Subjective=round(rowMeans(PostIntervention[SubjectiveVec],na.rm=TRUE),
1),

Behavior=round(rowMeans(PostIntervention[BehaviorVec],na.rm=TRUE),1))
FollowIntervention <-
transform(FollowIntervention,Declarative=round(rowMeans(FollowInterven
tion[DeclarativeVec],na.rm=TRUE),1),

Procedural=round(rowMeans(FollowIntervention[ProceduralVec],na.rm=TRUE
),1),
```

```
Subjective=round(rowMeans(FollowIntervention[SubjectiveVec],na.rm=TRUE
),1),

Behavior=round(rowMeans(FollowIntervention[BehaviorVec],na.rm=TRUE),1)
)
NonIntervention <-
transform(NonIntervention,Declarative=round(rowMeans(NonIntervention[D
eclarativeVec],na.rm=TRUE),1),

Procedural=round(rowMeans(NonIntervention[ProceduralVec],na.rm=TRUE),1
),

Subjective=round(rowMeans(NonIntervention[SubjectiveVec],na.rm=TRUE),1
),

Behavior=round(rowMeans(NonIntervention[BehaviorVec],na.rm=TRUE),1))

#Add variable for survey type
PreIntervention$Survey <-"Pre"
PostIntervention$Survey <- "Post"
FollowIntervention$Survey <- "Followup"
NonIntervention$Survey <- "Pre"

#We need to fix the factor levels so that all data frames have the
same set
Degree <- c("Administración Agropecuario","Desarrollo y Gestión
Interculturales","Economia Industrial","Fisioterapía","Odontología")
levels(PreIntervention$Degree) <-
c(levels(PreIntervention$Degree),"Economia Industrial")
levels(PostIntervention$Degree) <-
c(levels(PostIntervention$Degree),"Economia Industrial","Desarrollo y
Gestión Interculturales")
levels(FollowIntervention$Degree) <-
c(levels(FollowIntervention$Degree),"Economia Industrial","Desarrollo
y Gestión Interculturales")
levels(NonIntervention$Degree) <-
c(levels(NonIntervention$Degree),"Fisioterapía","Odontología")

PreIntervention$Degree <- factor(PreIntervention$Degree,levels=Degree)
PostIntervention$Degree <-
factor(PostIntervention$Degree,levels=Degree)
FollowIntervention$Degree <-
factor(FollowIntervention$Degree,levels=Degree)
FollowIntervention$Degree <-
factor(FollowIntervention$Degree,levels=Degree)

#Make Leon the base factor case
Origin <- c("Other","Leon","GTO","DF")
PreIntervention$Origin <- factor(PreIntervention$Origin,levels=Origin)
PostIntervention$Origin <-
factor(PostIntervention$Origin,levels=Origin)
```

```
FollowIntervention$Origin <-
factor(FollowIntervention$Origin,levels=Origin)
NonIntervention$Origin <- factor(NonIntervention$Origin,levels=Origin)

#Make Males the base factor
PreIntervention$Sex <- factor(PreIntervention$Sex,levels=c("m","f"))
PostIntervention$Sex <- factor(PostIntervention$Sex,levels=c("m","f"))
FollowIntervention$Sex <-
factor(FollowIntervention$Sex,levels=c("m","f"))
NonIntervention$Sex <- factor(NonIntervention$Sex,levels=c("m","f"))

#Create a new variable which groups the ages
PreIntervention$AgeCategory <- cut(PreIntervention$Age,c(0,23,30,100))
PostIntervention$AgeCategory <-
cut(PostIntervention$Age,c(0,23,30,100))
FollowIntervention$AgeCategory <-
cut(FollowIntervention$Age,c(0,23,30,100))
NonIntervention$AgeCategory <- cut(NonIntervention$Age,c(0,23,30,100))

#Instead We'll just change the failed grades from 0 to 5 so they don't
skew things quite so much
PreIntervention <- PreIntervention %>%
mutate(Final.Grade=replace(Final.Grade,Final.Grade==0,5))
PostIntervention <- PostIntervention %>%
mutate(Final.Grade=replace(Final.Grade,Final.Grade==0,5))
FollowIntervention <- FollowIntervention %>%
mutate(Final.Grade=replace(Final.Grade,Final.Grade==0,5))
NonIntervention <- NonIntervention %>%
mutate(Final.Grade=replace(Final.Grade,Final.Grade==0,5))

#Removing outliers using Cook's Distance and boxplots
PreIntervention <- PreIntervention[-c(27,78),]
PostIntervention <- PostIntervention[-c(45),]
FollowIntervention <- FollowIntervention[-c(21),]
NonIntervention <- NonIntervention[-c(17,26),]


#Create dataframe with the pre- observations from all of the students
Allstudentsurvey <- rbind(PreIntervention,NonIntervention)

#Create matching pre/post/followup dataframes
#Pre/Post
Presurvey_pre_postmatch <-
semi_join(PreIntervention,PostIntervention,by="UniqueID")
Postsurvey_pre_postmatch <-
semi_join(PostIntervention,PreIntervention,by="UniqueID")
Presurvey_pre_postmatch <- arrange(Presurvey_pre_postmatch,UniqueID)
Postsurvey_pre_postmatch <- arrange(Postsurvey_pre_postmatch,UniqueID)

#Pre/Followup
Presurvey_pre_followupmatch <-
semi_join(PreIntervention,FollowIntervention,by="UniqueID")
```

```
Followupsurvey_pre_followupmatch <-
semi_join(FollowIntervention,PreIntervention,by="UniqueID")
Presurvey_pre_followupmatch <-
arrange(Presurvey_pre_followupmatch,UniqueID)
Followupsurvey_pre_followupmatch<-
arrange(Followupsurvey_pre_followupmatch,UniqueID)

#Post/Followup
Postsurvey_post_followupmatch <-
semi_join(PostIntervention,FollowIntervention,by="UniqueID")
Followupsurvey_post_followupmatch <-
semi_join(FollowIntervention,PostIntervention,by="UniqueID")
Postsurvey_post_followupmatch <-
arrange(Postsurvey_post_followupmatch,UniqueID)
Followupsurvey_post_followupmatch<-
arrange(Followupsurvey_post_followupmatch,UniqueID)

#For Hypothesis 3 and 4 we are going to create dataframes of the
differences of individuals
Differences_Pre_Post <- Postsurvey_pre_postmatch[,8:33]-
Presurvey_pre_postmatch[,8:33]
Differences_Pre_Post <-
cbind(Postsurvey_pre_postmatch[,c(1:7,35)],Differences_Pre_Post)
Differences_Pre_Followup <- Followupsurvey_pre_followupmatch[,8:33]-
Presurvey_pre_followupmatch[,8:33]
Differences_Pre_Followup <-
cbind(Followupsurvey_pre_followupmatch[,c(1:7,35)],Differences_Pre_Fol
lowup)
Differences_Post_Followup <- Postsurvey_post_followupmatch[,8:33]-
Followupsurvey_post_followupmatch[,8:33]
Differences_Post_Followup <-
cbind(Followupsurvey_post_followupmatch[,c(1:7,35)],Differences_Post_F
ollowup)
```

E3. Identify and remove outliers

```
#Doing Regressions on all the datasets to look at relationship of
Behavior to predictors
RegressAll <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Allstudentsurvey)
RegressPre <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PreIntervention)
RegressPost <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PostIntervention)
RegressFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = FollowIntervention)
```

```
RegressDiff <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Post)
RegressDiff_PreFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Followup)

#ID outliers
Allcooks <- data.frame(cooks.distance(RegressAll))
boxplot(Allcooks)
Precooks<- data.frame(cooks.distance(RegressPre))
boxplot(Precooks)
Postcooks<- data.frame(cooks.distance(RegressPost))
boxplot(Postcooks)
Followcooks<- data.frame(cooks.distance(RegressFollow))
boxplot(Followcooks)
Diffcooks <- data.frame(cooks.distance(RegressDiff))
boxplot(Diffcooks)

#Saving a JPEG of the boxplot included in the Appendix
jpeg("Rresults/AllStudentCooksBoxplot.jpg")
boxplot(Allcooks)
dev.off()

#The outliers we chose to remove
PreIntervention <- PreIntervention[-c(27,78),]
PostIntervention <- PostIntervention[-c(45),]
FollowIntervention <- FollowIntervention[-c(21),]
Allstudentsurvey <- Allstudentsurvey[-c(27,78,101,110),]
```

E4. Produce descriptive statistics of the samples

```
#We are going to create a couple of custom function to produce a
simple table of descriptive statistics
#Reports the precentages of variables in one column
Percentage <- function(x){
  tbl <- round(prop.table(table(x))*100,digits=1)
  res <- cbind(tbl)
  colnames(res) <- c("x")
  res
}

#Takes the mean of a variable
MyMean <- function(x){
  m <- round(mean(x,na.rm=TRUE),1)
  m
}

#Creates a single column of all the variables we want in our
descriptive table for a dataset
DescriptiveStats <- function(x){
```

```
  Prec <- do.call(rbind,lapply(x[c(2,4,5)],Percentage))
  full <- rbind(nrow(x),Prec,MyMean(x$Age),MyMean(x$Final.Grade))
  full
  }


#Creating the descriptive stats for each dataset
All <- DescriptiveStats(Allstudentsurvey)
Intervention <- DescriptiveStats(PreIntervention)
Followup <- DescriptiveStats(FollowIntervention)

#Creating one table which will be outputed as an html table
DescriptiveTable <- cbind(All,Intervention,Followup)
row.names(DescriptiveTable) <-
c("Number","Male","Female","Elsewhere","Leon","Guanajuato
State","Mexico City","Agricultural Administration","Intercultural
Development and Management","Industrial Economics","Physical
Therapy","Dentistry","Age(mean)","Grade(mean)")
colnames(DescriptiveTable) <- c("All","Intervention","Follow-up")
stargazer(DescriptiveTable,summary=FALSE,out="Rresults/DescriptiveStat
s.html",digits=1, title = "Descriptive Statistics of the Surveyed
Samples")
```

E5. Compare the samples statistically

```
#First we will compare the samples of students who participated in the
intervention and thos we didn't by using their scores on the pre-class
survey
#We are going to look at the index scores (Declarative, Procedural,
Subjective and Behavior)
#First we'll take the means of each sample
Intermeans <-
round(do.call(rbind,lapply(PreIntervention[,30:33],mean)),2)
Nonmeans <-
round(do.call(rbind,lapply(NonIntervention[,30:33],mean)),2)

#The Kolmogorov-Smirnov test can be used to test the null-hypothesis
that two samples come from the same population based on a variable
#We will reject the null-hypothesis if the p-value is <0.05 at which
point the statistic estimates the size of the difference
KSstat <-
data.frame(t(mapply(ks.test,PreIntervention[,30:33],NonIntervention[,3
0:33])))
KSstat <- select(KSstat,statistic,p.value)
KSstat <-
transform(KSstat,statistic=round(as.numeric(statistic),2),p.value=roun
d(as.numeric(p.value),2))

#We combine these two comparison approaches into one table and print
it in HTML
IntervsNon<- data.frame(Intermeans,Nonmeans,KSstat)
```

```
colnames(IntervsNon) <- c("Intervention (mean)","Nonintervention
(mean)","KS Statistic","KS P-Value")
stargazer(IntervsNon,summary=FALSE,out="Rresults/InterVsNonSamples.htm
l",digits=2,title = "KS Test comparing Intervention and
Nonintervention Samples")

#Second we are concerned that there might be a statistical difference
in the students who bothered to follow-up one year later from the
overall intervention group
#So we will compare the pre-class scores of students who followed up
with those that didnt't
#For that we need to create two new databases
NonFollowInter <-
anti_join(PreIntervention,FollowIntervention,by="UniqueID")
FollowInterPre <-
semi_join(PreIntervention,FollowIntervention,by="UniqueID")

#Then we can repeate what we did for our first comparison
KSstat2 <-
data.frame(t(mapply(ks.test,FollowInterPre[,30:33],NonFollowInter[,30:
33])))
KSstat2 <- select(KSstat2,statistic,p.value)
KSstat2 <-
transform(KSstat2,statistic=round(as.numeric(statistic),2),p.value=rou
nd(as.numeric(p.value),2))
NonFollowmeans <-
round(do.call(rbind,lapply(NonFollowInter[,30:33],mean)),2)
FollowPremeans <-
round(do.call(rbind,lapply(FollowInterPre[,30:33],mean)),2)
FollowvsNon<- data.frame(FollowPremeans,NonFollowmeans,KSstat2)
colnames(FollowvsNon) <- c("Followups (mean)","Non-Followups
(mean)","KS Statistic","KS P-Value")

KSstat3 <-
ks.test(FollowInterPre$Final.Grade,NonFollowInter$Final.Grade)
NonFollowmeans3 <- round(mean(NonFollowInter$Final.Grade),2)
Followmeans3 <- round(mean(FollowInterPre$Final.Grade),2)
GradeFollow <-
data.frame(Followmeans3,NonFollowmeans3,KSstat3$statistic,KSstat3$p.va
lue)
colnames(GradeFollow) <- c("Followups (mean)","Non-Followups
(mean)","KS Statistic","KS P-Value")
row.names(GradeFollow) <- "Final Grade"

FollowvsNon <- rbind(FollowvsNon,GradeFollow)

stargazer(FollowvsNon,summary=FALSE,out="Rresults/FollowVsNonSamples.h
tml",digits=2,title = "KS Test comparing Preintervention sample with
Followup")
```

E6. Assessing the indices with Cronbach's Alpha

```
#Going to take a look at the indexes of the knowledge domains and
behavior for the all student survey
#First we'll calculate the mean and standard deviation for each
Allmeans <-
round(do.call(rbind,lapply(Allstudentsurvey[,30:33],mean)),2)
Allsd <- round(do.call(rbind,lapply(Allstudentsurvey[,30:33],sd)),2)

#To assess the index's reliability of measure we calculate Cronbach's
Alpha and report the standardized number (1 being all items are 100%
correlated)
DecA <- alpha(Allstudentsurvey[,DeclarativeVec])
DecP <- alpha(Allstudentsurvey[,ProceduralVec])
DecS <- alpha(Allstudentsurvey[,SubjectiveVec])
DecB <- alpha(Allstudentsurvey[,BehaviorVec])
Allalpha <-
t(data.frame(DecA$total[2],DecP$total[2],DecS$total[2],DecB$total[2]))
Allalpha <- transform(Allalpha,Alpha=round(as.numeric(X_data),2))
Allalpha <- select(Allalpha,Alpha)

#Create the html table
DomainsAll <- data.frame(Allmeans,Allsd,Allalpha)
colnames(DomainsAll) <- c("Mean","Standard Deviation", "Standardized
Alpha")
stargazer(DomainsAll,summary=FALSE,out="Rresults/DomainsAllAlpha.html"
,digits=2)

#Pre
Allmeans <-
round(do.call(rbind,lapply(PreIntervention[,30:33],mean)),2)
Allsd <- round(do.call(rbind,lapply(PreIntervention[,30:33],sd)),2)
DecA <- alpha(PreIntervention[,DeclarativeVec])
DecP <- alpha(PreIntervention[,ProceduralVec])
DecS <- alpha(PreIntervention[,SubjectiveVec])
DecB <- alpha(PreIntervention[,BehaviorVec])
Allalpha <-
t(data.frame(DecA$total[2],DecP$total[2],DecS$total[2],DecB$total[2]))
Allalpha <- transform(Allalpha,Alpha=round(as.numeric(X_data),2))
Allalpha <- select(Allalpha,Alpha)
DomainsPre <- data.frame(Allmeans,Allsd,Allalpha)
colnames(DomainsPre) <- c("Mean","Standard Deviation", "Standardized
Alpha")
stargazer(DomainsPre,summary=FALSE,out="Rresults/DomainsPreAlpha.html"
,digits=2)

#Post
Allmeans <-
round(do.call(rbind,lapply(PostIntervention[,30:33],mean)),2)
Allsd <- round(do.call(rbind,lapply(PostIntervention[,30:33],sd)),2)
DecA <- alpha(PostIntervention[,DeclarativeVec])
```

```
DecP <- alpha(PostIntervention[,ProceduralVec])
DecS <- alpha(PostIntervention[,SubjectiveVec])
DecB <- alpha(PostIntervention[,BehaviorVec])
Allalpha <-
t(data.frame(DecA$total[2],DecP$total[2],DecS$total[2],DecB$total[2]))
Allalpha <- transform(Allalpha,Alpha=round(as.numeric(X_data),2))
Allalpha <- select(Allalpha,Alpha)
DomainsPost <- data.frame(Allmeans,Allsd,Allalpha)
colnames(DomainsPost) <- c("Mean","Standard Deviation", "Standardized
Alpha")
stargazer(DomainsPost,summary=FALSE,out="Rresults/DomainsPostAlpha.htm
l",digits=2)

#Follow
Allmeans <-
round(do.call(rbind,lapply(FollowIntervention[,30:33],mean)),2)
Allsd <- round(do.call(rbind,lapply(FollowIntervention[,30:33],sd)),2)
DecA <- alpha(FollowIntervention[,DeclarativeVec])
DecP <- alpha(FollowIntervention[,ProceduralVec])
DecS <- alpha(FollowIntervention[,SubjectiveVec])
DecB <- alpha(FollowIntervention[,BehaviorVec])
Allalpha <-
t(data.frame(DecA$total[2],DecP$total[2],DecS$total[2],DecB$total[2]))
Allalpha <- transform(Allalpha,Alpha=round(as.numeric(X_data),2))
Allalpha <- select(Allalpha,Alpha)
DomainsFollow <- data.frame(Allmeans,Allsd,Allalpha)
colnames(DomainsFollow) <- c("Mean","Standard Deviation",
"Standardized Alpha")
stargazer(DomainsFollow,summary=FALSE,out="Rresults/DomainsFollowAlpha
.html",digits=2)
```

E7. Multicollinearity among independent variables in the regressions

```
RegressAll <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Allstudentsurvey)
RegressPre <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PreIntervention)
RegressPost <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PostIntervention)
RegressFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = FollowIntervention)
RegressDiff <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Post)
RegressDiff_PreFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Followup)
```

```
#Check for colinearity among regressors by calculating Variance
Influence Factors
VarianceInfluenceFactors <-
data.frame(vif(RegressAll),vif(RegressPre),vif(RegressPost),vif(Regres
sFollow),vif(RegressDiff),vif(RegressDiff_PreFollow))
VarianceInfluenceFactors <-
dplyr::select(VarianceInfluenceFactors,GVIF,GVIF.1,GVIF.2,GVIF.3,GVIF.
4,GVIF.5)
colnames(VarianceInfluenceFactors) <- c("All Students","Pre
Intervention","Post Intervention","Followup","Differences
Pre/Post","Differences Pre/Followup")
stargazer(VarianceInfluenceFactors,summary=FALSE,
out="Rresults/VIF.html",title = "Variance Influence Factors for all
Regressions")

#The VIFs are all low except for the Follow-up sample so we need to be
very cautious about interpreting those results
#For the follow-up sample the three knowledge indices appear to be
tightly correlated with each other
```

E8. Table of All Regression Results

```
RegressAll <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Allstudentsurvey)
StdRegressAll <- standardize(RegressAll)
RegressPre <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PreIntervention)
StdRegressPre <- standardize(RegressPre)
RegressPost <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PostIntervention)
StdRegressPost <- standardize(RegressPost)
RegressFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = FollowIntervention)
StdRegressFollow <- standardize(RegressFollow)
RegressDiff <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Post)
StdRegressDiff <- standardize(RegressDiff)
RegressDiff_PreFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Followup)
StdRegressDiff_PreFollow <- standardize(RegressDiff_PreFollow)

stargazer(RegressAll,RegressPre,RegressPost,RegressFollow,RegressDiff,
RegressDiff_PreFollow, out = "Rresults/AllRegressions.html",
column.labels = c("All Students","Pre Intervention","Post
```

```
Intervention","Followup","Differences Pre/Post","Differences
Pre/Followup"),title = "Summary of all regression results in study")
stargazer(StdRegressAll,StdRegressPre,StdRegressPost,StdRegressFollow,
StdRegressDiff,StdRegressDiff_PreFollow, out =
"Rresults/AllRegressionsStandardized.html", column.labels = c("All
Students","Pre Intervention","Post
Intervention","Followup","Differences Pre/Post","Differences
Pre/Followup"),title = "Summary of standardized regression results in
study")
```

E9. Evaluating Hypothesis 1

```
#Do regression on dataset of all-students to look at relationship of
Behavior and Indices
RegressAll <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Allstudentsurvey)
StdRegressAll <- (RegressAll)
stargazer(StdRegressAll,out =
"Rresults/H1_AllStudentSurveyRegression.html", column.labels =
"Standardized All Students",title = "Hypothesis 1: Regression
Examining Relationship between Knowledge and Behavior")
```

E10. Evaluating Hypothesis 2

```
#We conduct a paired T-Test comparing each students pre and post
intervention survey responses
Ttest_PrePostQs <-
mapply(t.test,Postsurvey_pre_postmatch[,8:29],Presurvey_pre_postmatch[
,8:29],paired=TRUE, conf.level=0.95)
Ttest_PrePostQs <- data.frame(Ttest_PrePostQs)
Ttest_PrePostQs <- data.frame(t(Ttest_PrePostQs))
Ttest_PrePostQs[,c(6,7,8,9)] <- list(NULL)
Ttest_PrePostQs <- cSplit(Ttest_PrePostQs,"conf.int")
Ttest_PrePostQs <-
transform(Ttest_PrePostQs,Tstatistic=as.numeric(statistic),
DegreesFreedom=as.numeric(parameter),
p.value=as.numeric(p.value),estimate=as.numeric(estimate),

conf.int_min=extract_numeric(conf.int_1),conf.int_max=extract_numeric(
conf.int_2))
Ttest_PrePostQs[,c(1,2,5,6)] <- list(NULL)
#Adjust p-values for multiple comparisons using Benjamini & Yekkutieli
(2001)
Ttest_PrePostQs <-
mutate(Ttest_PrePostQs,Adjusted.P.Values=p.adjust(p.value,method="BY")
)
Ttest_PrePostQs$Names <- xaxis_names[1:22]
Ttest_PrePostQs$Names <-
factor(Ttest_PrePostQs$Names,levels=Ttest_PrePostQs$Names)
#Effect Size with Cohen's D
```

```
CohensDEffect_PrePostQs <-
mapply(cohen.d,Postsurvey_pre_postmatch[,8:29],Presurvey_pre_postmatch
[,8:29],paired=TRUE)
CohensDEffect_PrePostQs <- data.frame(CohensDEffect_PrePostQs)
CohensDEffect_PrePostQs <- data.frame(t(CohensDEffect_PrePostQs))
CohensDEffect_PrePostQs[,c(1,2,5,6)] <- list(NULL)
CohensDEffect_PrePostQs <- cSplit(CohensDEffect_PrePostQs,"conf.int")
CohensDEffect_PrePostQs <-
transform(CohensDEffect_PrePostQs,CohensD=as.numeric(estimate),CD.conf
.int_min=extract_numeric(conf.int_1),CD.conf.int_max=extract_numeric(c
onf.int_2))
CohensDEffect_PrePostQs[,c(1,2,3)] <- list(NULL)
TtestCohensD_PrePostQs <-
data.frame(Ttest_PrePostQs,CohensDEffect_PrePostQs)
TtestCohensD_PrePostQs <-
TtestCohensD_PrePostQs[,c(8,2,1,7,5,6,3,4,9,10,11)]
TtestCohensD_PrePostQs$Names <-
factor(TtestCohensD_PrePostQs$Names,levels=TtestCohensD_PrePostQs$Name
s)
#Printing the Results to HTML
stargazer(TtestCohensD_PrePostQs,title="T-test Results and Cohen's D
for Pre vs. Post Change in Survey",
summary=FALSE,out="Rresults/H2_PrePostTtestAllQs.html")
#Graphing the results and saving
ggplot(TtestCohensD_PrePostQs, aes(x=Names, y=estimate,
ymin=conf.int_min,ymax=conf.int_max),ordered=FALSE) +
geom_pointrange() + theme_bw() + coord_flip() + geom_hline(yintercept
= 0,linetype = "dotted") + ylab("T test Estimate of Mean Response
Change")+xlab("")
ggsave("Rresults/H2_GraphPrePostAllQs_Vertical.jpg")


#Ttest for just the indices which is another way of dealing with
multiple comparisons
Ttest_PrePost <-
mapply(t.test,Postsurvey_pre_postmatch[,30:33],Presurvey_pre_postmatch
[,30:33],paired=TRUE, conf.level=0.95)
Ttest_PrePost <- data.frame(Ttest_PrePost)
Ttest_PrePost <- data.frame(t(Ttest_PrePost))
Ttest_PrePost[,c(6,7,8,9)] <- list(NULL)
Ttest_PrePost <- cSplit(Ttest_PrePost,"conf.int")
Ttest_PrePost <-
transform(Ttest_PrePost,Tstatistic=as.numeric(statistic),
DegreesFreedom=as.numeric(parameter),
p.value=as.numeric(p.value),estimate=as.numeric(estimate),

conf.int_min=extract_numeric(conf.int_1),conf.int_max=extract_numeric(
conf.int_2))
Ttest_PrePost[,c(1,2,5,6)] <- list(NULL)
#Adjust p-values for multiple comparisons using Benjamini & Yekkutieli
(2001)
```

```
Ttest_PrePost <-
mutate(Ttest_PrePost,Adjusted.P.Values=p.adjust(p.value,method="BY"))
Ttest_PrePost$Names <-
c("Declarative","Procedural","Subjective","Behavior")
Ttest_PrePost$Names <-
factor(Ttest_PrePost$Names,levels=Ttest_PrePost$Names)
#Effect Size with Cohen's D
CohensDEffect_PrePost <-
mapply(cohen.d,Postsurvey_pre_postmatch[,30:33],Presurvey_pre_postmatc
h[,30:33],paired=TRUE)
CohensDEffect_PrePost <- data.frame(CohensDEffect_PrePost)
CohensDEffect_PrePost <- data.frame(t(CohensDEffect_PrePost))
CohensDEffect_PrePost[,c(1,2,5,6)] <- list(NULL)
CohensDEffect_PrePost <- cSplit(CohensDEffect_PrePost,"conf.int")
CohensDEffect_PrePost <-
transform(CohensDEffect_PrePost,CohensD=as.numeric(estimate),CD.conf.i
nt_min=extract_numeric(conf.int_1),CD.conf.int_max=extract_numeric(con
f.int_2))
CohensDEffect_PrePost[,c(1,2,3)] <- list(NULL)
TtestCohensD_PrePostQs <-
data.frame(Ttest_PrePost,CohensDEffect_PrePost)
TtestCohensD_PrePostQs <-
TtestCohensD_PrePostQs[,c(8,2,1,7,5,6,3,4,9,10,11)]
TtestCohensD_PrePostQs$Names <-
factor(TtestCohensD_PrePostQs$Names,levels=TtestCohensD_PrePostQs$Name
s)
#Printing the Results to HTML
stargazer(TtestCohensD_PrePostQs,title="T-test Results and Cohen's D
for Pre vs. Post Change in Indices",
summary=FALSE,out="Rresults/H2_PrePostTtestIndices.html")
#Graphing the results and saving
ggplot(TtestCohensD_PrePostQs, aes(x=Names, y=CohensD,
ymin=CD.conf.int_min,ymax=CD.conf.int_max),ordered=FALSE) +
geom_pointrange(size=2,shape=15) +
theme_bw()+theme(text=element_text(size=20)) + coord_flip() +
geom_hline(yintercept = 0,linetype = "longdash",size=1) + ylab(""Cohen
D Effect Size Estimate"")+xlab("")
ggsave("Rresults/H2_GraphPrePostIndices_Vertical.jpg")
```

E11. Evaluating Hypothesis 3

```
#Data frame created with the differences between pre and post for
individuals
#Run the same regression as for H1
RegressDiff <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Post)
STDRegressDiff <- standardize(RegressDiff)
stargazer(STDRegressDiff,out="Rresults/H3_RegressionOfDifferences.html
",column.labels = "Difference Between Pre/Post",title = "Hypothesis 3:
Relationship between the change in knowledge and change in behavior")
```

E12. Evaluating Hypothesis 4

```
#Main test of hypothesis is to see if t-test shows difference between
pre and followup
Ttest_PreFollowQs <-
mapply(t.test,Followupsurvey_pre_followupmatch[,8:29],Presurvey_pre_fo
llowupmatch[,8:29],paired=TRUE, conf.level=0.95)
Ttest_PreFollowQs <- data.frame(Ttest_PreFollowQs)
Ttest_PreFollowQs <- data.frame(t(Ttest_PreFollowQs))
Ttest_PreFollowQs[,c(6,7,8,9)] <- list(NULL)
Ttest_PreFollowQs <- cSplit(Ttest_PreFollowQs,"conf.int")
Ttest_PreFollowQs <-
transform(Ttest_PreFollowQs,Tstatistic=as.numeric(statistic),
DegreesFreedom=as.numeric(parameter),
p.value=as.numeric(p.value),estimate=as.numeric(estimate),

conf.int_min=extract_numeric(conf.int_1),conf.int_max=extract_numeric(
conf.int_2))
Ttest_PreFollowQs[,c(1,2,5,6)] <- list(NULL)
#Adjust p-values for multiple comparisons using Benjamini & Yekkutieli
(2001)
Ttest_PreFollowQs <-
mutate(Ttest_PreFollowQs,Adjusted.P.Values=p.adjust(p.value,method="BY
"))
Ttest_PreFollowQs$Names <- xaxis_names[1:22]
Ttest_PreFollowQs$Names <-
factor(Ttest_PreFollowQs$Names,levels=Ttest_PreFollowQs$Names)
#Effect Size with Cohen's D
CohensDEffect_PreFollowQs <-
mapply(cohen.d,Followupsurvey_pre_followupmatch[,8:29],Presurvey_pre_f
ollowupmatch[,8:29],paired=TRUE)
CohensDEffect_PreFollowQs <- data.frame(CohensDEffect_PreFollowQs)
CohensDEffect_PreFollowQs <- data.frame(t(CohensDEffect_PreFollowQs))
CohensDEffect_PreFollowQs[,c(1,2,5,6)] <- list(NULL)
CohensDEffect_PreFollowQs <-
cSplit(CohensDEffect_PreFollowQs,"conf.int")
CohensDEffect_PreFollowQs <-
transform(CohensDEffect_PreFollowQs,CohensD=as.numeric(estimate),CD.co
nf.int_min=extract_numeric(conf.int_1),CD.conf.int_max=extract_numeric
(conf.int_2))
CohensDEffect_PreFollowQs[,c(1,2,3)] <- list(NULL)
TtestCohensD_PreFollowQs <-
data.frame(Ttest_PreFollowQs,CohensDEffect_PreFollowQs)
TtestCohensD_PreFollowQs <-
TtestCohensD_PreFollowQs[,c(8,2,1,7,5,6,3,4,9,10,11)]
TtestCohensD_PreFollowQs$Names <-
factor(TtestCohensD_PreFollowQs$Names,levels=TtestCohensD_PreFollowQs$
Names)
#Printing the Results to HTML
```

```
stargazer(TtestCohensD_PreFollowQs,title="T-test Results and Cohen's D
for Pre vs. Follow Change in Survey",
summary=FALSE,out="Rresults/H4_PreFollowTtestAllQs.html")
#Graphing the results and saving
ggplot(TtestCohensD_PreFollowQs, aes(x=Names, y=estimate,
ymin=conf.int_min,ymax=conf.int_max),ordered=FALSE) +
geom_pointrange() + theme_bw() + coord_flip() + geom_hline(yintercept
= 0,linetype = "dotted") + ylab("T test Estimate of Mean Response
Change")+xlab("")
ggsave("Rresults/H4_GraphPreFollowAllQs_Vertical.jpg")


#Ttest for just the indices which is another way of dealing with
multiple comparisons
Ttest_PreFollow <-
mapply(t.test,Followupsurvey_pre_followupmatch[,30:33],Presurvey_pre_f
ollowupmatch[,30:33],paired=TRUE, conf.level=0.95)
Ttest_PreFollow <- data.frame(Ttest_PreFollow)
Ttest_PreFollow <- data.frame(t(Ttest_PreFollow))
Ttest_PreFollow[,c(6,7,8,9)] <- list(NULL)
Ttest_PreFollow <- cSplit(Ttest_PreFollow,"conf.int")
Ttest_PreFollow <-
transform(Ttest_PreFollow,Tstatistic=as.numeric(statistic),
DegreesFreedom=as.numeric(parameter),
p.value=as.numeric(p.value),estimate=as.numeric(estimate),

conf.int_min=extract_numeric(conf.int_1),conf.int_max=extract_numeric(
conf.int_2))
Ttest_PreFollow[,c(1,2,5,6)] <- list(NULL)
#Adjust p-values for multiple comparisons using Benjamini & Yekkutieli
(2001)
Ttest_PreFollow <-
mutate(Ttest_PreFollow,Adjusted.P.Values=p.adjust(p.value,method="BY")
)
Ttest_PreFollow$Names <-
c("Declarative","Procedural","Subjective","Behavior")
Ttest_PreFollow$Names <-
factor(Ttest_PreFollow$Names,levels=Ttest_PreFollow$Names)
#Effect Size with Cohen's D
CohensDEffect_PreFollow <-
mapply(cohen.d,Followupsurvey_pre_followupmatch[,30:33],Presurvey_pre_
followupmatch[,30:33],paired=TRUE)
CohensDEffect_PreFollow <- data.frame(CohensDEffect_PreFollow)
CohensDEffect_PreFollow <- data.frame(t(CohensDEffect_PreFollow))
CohensDEffect_PreFollow[,c(1,2,5,6)] <- list(NULL)
CohensDEffect_PreFollow <- cSplit(CohensDEffect_PreFollow,"conf.int")
CohensDEffect_PreFollow <-
transform(CohensDEffect_PreFollow,CohensD=as.numeric(estimate),CD.conf
.int_min=extract_numeric(conf.int_1),CD.conf.int_max=extract_numeric(c
onf.int_2))
CohensDEffect_PreFollow[,c(1,2,3)] <- list(NULL)
```

```
TtestCohensD_PreFollow <-
data.frame(Ttest_PreFollow,CohensDEffect_PreFollow)
TtestCohensD_PreFollow <-
TtestCohensD_PreFollow[,c(8,2,1,7,5,6,3,4,9,10,11)]
TtestCohensD_PreFollow$Names <-
factor(TtestCohensD_PreFollow$Names,levels=TtestCohensD_PreFollow$Name
s)
#Printing the Results to HTML
stargazer(TtestCohensD_PreFollow,title="T-test Results and Cohen's D
for Pre vs. Followup Change in Indices",
summary=FALSE,out="Rresults/H4_PreFollowupTtestIndices.html")
#Graphing the results and saving
ggplot(TtestCohensD_PreFollow, aes(x=Names, y=CohensD,
ymin=CD.conf.int_min,ymax=CD.conf.int_max),ordered=FALSE) +
geom_pointrange(size=2,shape=15) +
theme_bw()+theme(text=element_text(size=20)) + coord_flip() +
geom_hline(yintercept = 0,linetype = "longdash",size=1) + ylab("Cohen
D Effect Size Estimate")+xlab("")
ggsave("Rresults/H4_GraphPreFollowIndices_Vertical.jpg")
```

E13. Density Plot Histograms of the Indices

```
#Denisty Plot Histograms of indexes
Alldata <- rbind(Allstudentsurvey,PostIntervention,FollowIntervention)
Alldata$Survey <-
factor(Alldata$Survey,levels=c("Pre","Post","Followup"))

ggplot(Alldata,aes(x=Behavior,
fill=Survey))+geom_density(alpha=.3)+xlim(1,4)+ylim(0,1.3)+ylab("Frequ
ency of Index
Scores")+theme(text=element_text(size=20),legend.position="bottom",leg
end.title=element_blank())
ggsave("Rresults/DensityPlotBehavior.jpeg")
ggplot(Alldata,aes(x=Declarative,
fill=Survey))+geom_density(alpha=.3)+xlim(1,4)+ylim(0,1.3)+ylab("Frequ
ency of Index
Scores")+theme(text=element_text(size=20),legend.position="bottom",leg
end.title=element_blank())
ggsave("Rresults/DensityPlotDeclarative.jpeg")
ggplot(Alldata,aes(x=Procedural,
fill=Survey))+geom_density(alpha=.3)+xlim(1,4)+ylim(0,1.3)+ylab("Frequ
ency of Index
Scores")+theme(text=element_text(size=20),legend.position="bottom",leg
end.title=element_blank())
ggsave("Rresults/DensityPlotProcedural.jpeg")
ggplot(Alldata,aes(x=Subjective,
fill=Survey))+geom_density(alpha=.3)+xlim(1,4)+ylim(0,1.3)+ylab("Frequ
ency of Index
Scores")+theme(text=element_text(size=20),legend.position="bottom",leg
end.title=element_blank())
ggsave("Rresults/DensityPlotSubjective.jpeg")
```

E14. Investigation into the Sex Differences

```
#First we are create new databases of just the female and just the
male students
All_Female <- Allstudentsurvey %>% filter(Sex=="f")
All_Male <- Allstudentsurvey %>% filter(Sex=="m")
Post_Female <- PostIntervention %>% filter(Sex=="f")
Post_Male <- PostIntervention %>% filter(Sex=="m")
Follow_Female <- FollowIntervention %>% filter(Sex=="f")
Follow_Male <- FollowIntervention %>% filter(Sex=="m")

#Ran OLs regressions on each sex
RegressAll_Female <-
lm(Behavior~AgeCategory+Origin+Degree+Final.Grade+Declarative+Procedur
al+Subjective,data = All_Female)
RegressAll_Male <-
lm(Behavior~AgeCategory+Origin+Degree+Final.Grade+Declarative+Procedur
al+Subjective,data = All_Male)
stargazer(RegressAll_Female,RegressAll_Male,
out="Rresults/SexRegressions.html", column.labels =
c("Female","Male"), title = "Regressions of All Student Survey
Seperated by Sex")

STDRegressAll_Female <- standardize(RegressAll_Female)
STDRegressAll_Male <- standardize(RegressAll_Male)
stargazer(STDRegressAll_Female,STDRegressAll_Male,
out="Rresults/SexRegressionsSTD.html", column.labels =
c("Female","Male"), title = "Standardized Regressions of All Student
Survey Seperated by Sex")

#Ttests comparing male and female sustainable behaviors for each of
the three samples
#Then lots of work to get it in a usable format
Ttest_Sex_All <- matrix(t.test(All_Female$Behavior,All_Male$Behavior))
Ttest_Sex_Post <-
matrix(t.test(Post_Female$Behavior,Post_Male$Behavior))
Ttest_Sex_Follow <-
matrix(t.test(Follow_Female$Behavior,Follow_Male$Behavior))
Ttest_Sex <- data.frame(Ttest_Sex_All,Ttest_Sex_Post,Ttest_Sex_Follow)
Ttest_Sex <- data.frame(t(Ttest_Sex))
Ttest_Sex[,c(6,7,8,9)] <- list(NULL)
Ttest_Sex <- cSplit(Ttest_Sex,c("X4","X5"))
Ttest_Sex <-
transmute(Ttest_Sex,t.statistic=as.numeric(X1),parameter=as.numeric(X2
),p.value=as.numeric(X3),FemaleMean=extract_numeric(X5_1),MaleMean=ext
ract_numeric(X5_2),DiffofMeans=FemaleMean-MaleMean,
confint_min=extract_numeric(X4_1),confint_max=extract_numeric(X4_2))
Ttest_Sex$Surveys <- c("All","Post","Followup")
Ttest_Sex$Surveys <-
factor(Ttest_Sex$Surveys,levels=Ttest_Sex$Surveys)
```

```
ggplot(Ttest_Sex, aes(x=Surveys, y=DiffofMeans,
ymin=confint_min,ymax=confint_max),ordered=FALSE) +
geom_pointrange(size=2,shape=15) +
theme_bw()+theme(text=element_text(size=20)) + coord_flip() +
geom_hline(yintercept = 0,linetype = "longdash",size=1) +
ylab("Difference Between the Sexes")+xlab("")
ggsave("Rresults/GraphofSexDiff.jpeg")
stargazer(Ttest_Sex,out="Rresults/Sexttest.html",summary=FALSE)

#Another way to visualize the difference:
ggplot(Allstudentsurvey,aes(x=Behavior, group=Sex,
fill=Sex))+geom_density(alpha=.3)+xlim(1,4)+ylab("Frequency of Index
Scores")+theme(text=element_text(size=20),legend.position="bottom",leg
end.title=element_blank())
ggsave("Rresults/DensityPlotSexAll.jpeg")

#KS Test:
KSstatSex <-
data.frame(t(mapply(ks.test,All_Female[,30:33],All_Male[,30:33])))
KSstatSex <- dplyr::select(KSstatSex,statistic,p.value)
KSstatSex <-
transform(KSstatSex,statistic=round(as.numeric(statistic),2),p.value=r
ound(as.numeric(p.value),2))
Femalemeans <- round(do.call(rbind,lapply(All_Female[,30:33],mean)),2)
Malemeans <- round(do.call(rbind,lapply(All_Male[,30:33],mean)),2)
FemaleVsMale<- data.frame(Femalemeans,Malemeans,KSstatSex)
colnames(FemaleVsMale) <- c("Female (mean)","Male (mean)","KS
Statistic","KS P-Value")
stargazer(FemaleVsMale,summary=FALSE,out="Rresults/KSSexComparison.htm
l",digits=2,title = "KS Test comparing Females vs Males for All
student Sample")
```