data analysis for the interaction between age group and word predictability only on Identical preview condition

2024-06-03

Version: Rtudio: 2023.12.0 Build 369 R:4.3.2 lme4:1.1-35.1 lattice:0.21-9 MASS:7.3-60 plyr:1.8.9

Load the packages that we used

rm(list=ls()) # clear objects from the workspace  
datafile <- read.csv("datafile.csv", sep = ",", dec = ".") # open the datafile  
colnames(datafile) #retrieve the column names of the datafile

## [1] "Participant" "Item"   
## [3] "Age" "Predictability"   
## [5] "Preview" "Age\_Predictability"   
## [7] "Age\_Preview" "Predictability\_Preview"   
## [9] "Condition" "TRT"   
## [11] "IA\_SKIP" "FFD"   
## [13] "SFD" "GD"   
## [15] "RO" "RI"   
## [17] "sentence.reading.time" "average.fixation.duration"   
## [19] "number.of.fixations" "number.of.regressions"   
## [21] "average.forwary.saccade.length"

nrow(datafile) #the number of observations of the datafile

## [1] 6811

datafile$pp = datafile[,1]  
datafile$Item = datafile[,2]  
datafile$Age = datafile[,3]  
datafile$Predictability = datafile[,4]  
datafile$Preview = datafile[,5]  
datafile$Age\_Pred = datafile[,6]  
datafile$Age\_Preview = datafile[,7]  
datafile$Pred\_Preview = datafile[,8]  
datafile$Condition = datafile[,9]  
  
# Let's make sure all the variables are from the correct class  
datafile$Item = as.factor(datafile$Item)  
datafile$Age = factor(datafile$Age, levels = c("Y","O"))  
datafile$Predictability = factor(datafile$Predictability, levels = c("H","L"))  
datafile$Preview = factor(datafile$Preview, levels = c("I","T","S"))  
datafile$Age\_Pred = factor(datafile$Age\_Pred, levels = c("YH","YL","OH","OL"))  
datafile$Age\_Preview = factor(datafile$Age\_Preview, levels = c("YI","YT","YS","OI","OT","OS"))  
datafile$Pred\_Preview = factor(datafile$Pred\_Preview, levels = c("HI","HT","HS","LI","LT","LS"))  
datafile$Condition = factor(datafile$Condition, levels = c("YHI","YHT","YHS","YLI","YLT","YLS","OHI","OHT","OHS","OLI","OLT","OLS"))

datafile<-subset(datafile, datafile$Preview=="I") #Identical preview  
nrow(datafile)

## [1] 2264

contrasts(datafile$Age) <- contr.sdif(2)   
contrasts(datafile$Predictability) <- contr.sdif(2)

# SKIPPING RATE

measure = "IA\_SKIP"   
datafile$depvar = datafile[,measure]  
datafile$depvar = as.numeric(datafile$depvar)

## linear mixed effect model analysis for skipping rate

sk\_model= glmer(depvar~Age\*Predictability + (1+Predictability|pp) + (1|Item),datafile,family = binomial)  
summary(sk\_model)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: depvar ~ Age \* Predictability + (1 + Predictability | pp) + (1 |   
## Item)  
## Data: datafile  
##   
## AIC BIC logLik deviance df.resid   
## 2566.8 2612.6 -1275.4 2550.8 2256   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.4004 -0.7055 -0.3486 0.9380 3.8017   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr  
## pp (Intercept) 0.415799 0.64482   
## Predictability2-1 0.004205 0.06485 1.00  
## Item (Intercept) 0.083501 0.28897   
## Number of obs: 2264, groups: pp, 96; Item, 90  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.13403 0.09762 -11.616 <2e-16 \*\*\*  
## Age2-1 -1.70529 0.18398 -9.269 <2e-16 \*\*\*  
## Predictability2-1 -0.26916 0.12161 -2.213 0.0269 \*   
## Age2-1:Predictability2-1 -0.38973 0.23801 -1.637 0.1015   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) Age2-1 Prd2-1  
## Age2-1 0.373   
## Prdctblt2-1 0.128 0.093   
## Ag2-1:Pr2-1 0.090 0.122 0.540

confint(sk\_model, method="Wald")

## 2.5 % 97.5 %  
## .sig01 NA NA  
## .sig02 NA NA  
## .sig03 NA NA  
## .sig04 NA NA  
## (Intercept) -1.3253626 -0.94268797  
## Age2-1 -2.0658842 -1.34469122  
## Predictability2-1 -0.5075008 -0.03081330  
## Age2-1:Predictability2-1 -0.8562159 0.07674927

## simple effect analysis for skipping rate

contrast.matrix <- matrix(c(  
 -1, 1, 0, 0,   
 0, 0, -1, 1), 4, 2,  
 dimnames=list(c("YH","YL","OH","OL"),   
 c("Y:L-H", "O:L-H")))  
inv.contrast.matrix <- matrix(t(ginv(contrast.matrix)), 4, 2,  
 dimnames=list(c("YH","YL","OH","OL"),   
 c("Y:L-H", "O:L-H")))  
(contrasts(datafile$Age\_Pred) <- fractions(inv.contrast.matrix))

## Y:L-H O:L-H  
## YH -1/2 0   
## YL 1/2 0   
## OH 0 -1/2   
## OL 0 1/2

depvar.lmer1 = glmer(depvar ~ Age\_Pred + (1|pp)+(1|Item), datafile,family=binomial)  
summary(depvar.lmer1, corr = FALSE)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: depvar ~ Age\_Pred + (1 | pp) + (1 | Item)  
## Data: datafile  
##   
## AIC BIC logLik deviance df.resid   
## 2563.0 2597.4 -1275.5 2551.0 2258   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.4051 -0.7093 -0.3488 0.9391 3.7357   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## pp (Intercept) 0.410 0.6403   
## Item (Intercept) 0.083 0.2881   
## Number of obs: 2264, groups: pp, 96; Item, 90  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.13173 0.09701 -11.666 <2e-16 \*\*\*  
## Age\_PredY:L-H -0.07047 0.11476 -0.614 0.5392   
## Age\_PredO:L-H -0.42752 0.19572 -2.184 0.0289 \*   
## Age\_Pred -1.70252 0.18295 -9.306 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# FIRST FIXATION DURATION (FFD)

measure = "FFD"   
datafile$depvar = log(datafile[,measure])  
datafile$depvar = as.numeric(datafile$depvar)

## linear mixed effect model analysis for FFD

FFD\_model= lmer(depvar~Age\*Predictability + (1+Predictability|pp) + (1+Predictability|Item),datafile)  
summary(FFD\_model)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: depvar ~ Age \* Predictability + (1 + Predictability | pp) + (1 +   
## Predictability | Item)  
## Data: datafile  
##   
## REML criterion at convergence: 594.8  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.4856 -0.5990 0.0053 0.5748 4.1135   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## pp (Intercept) 0.0100620 0.10031   
## Predictability2-1 0.0023492 0.04847 0.22   
## Item (Intercept) 0.0008645 0.02940   
## Predictability2-1 0.0018330 0.04281 -0.06  
## Residual 0.0777483 0.27883   
## Number of obs: 1540, groups: pp, 96; Item, 90  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 5.43668 0.01326 410.060  
## Age2-1 0.12432 0.02578 4.823  
## Predictability2-1 0.04443 0.01598 2.780  
## Age2-1:Predictability2-1 0.01066 0.03065 0.348  
##   
## Correlation of Fixed Effects:  
## (Intr) Age2-1 Prd2-1  
## Age2-1 0.155   
## Prdctblt2-1 0.041 0.010   
## Ag2-1:Pr2-1 0.010 0.048 0.016

confint(FFD\_model, method="Wald")

## 2.5 % 97.5 %  
## .sig01 NA NA  
## .sig02 NA NA  
## .sig03 NA NA  
## .sig04 NA NA  
## .sig05 NA NA  
## .sig06 NA NA  
## .sigma NA NA  
## (Intercept) 5.41069922 5.46267074  
## Age2-1 0.07379735 0.17484523  
## Predictability2-1 0.01310016 0.07575186  
## Age2-1:Predictability2-1 -0.04940562 0.07073380

## simple effect analysis for FFD

contrast.matrix <- matrix(c(  
 -1, 1, 0, 0,   
 0, 0, -1, 1), 4, 2,  
 dimnames=list(c("YH","YL","OH","OL"),   
 c("Y:L-H", "O:L-H")))  
inv.contrast.matrix <- matrix(t(ginv(contrast.matrix)), 4, 2,  
 dimnames=list(c("YH","YL","OH","OL"),   
 c("Y:L-H", "O:L-H")))  
(contrasts(datafile$Age\_Pred) <- fractions(inv.contrast.matrix))

## Y:L-H O:L-H  
## YH -1/2 0   
## YL 1/2 0   
## OH 0 -1/2   
## OL 0 1/2

depvar.lmer1 = lmer(depvar ~ Age\_Pred + (1|pp)+(1|Item), datafile)  
summary(depvar.lmer1, corr = FALSE)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: depvar ~ Age\_Pred + (1 | pp) + (1 | Item)  
## Data: datafile  
##   
## REML criterion at convergence: 596  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.4649 -0.6112 0.0080 0.5893 4.1961   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## pp (Intercept) 0.0100603 0.10030   
## Item (Intercept) 0.0009031 0.03005   
## Residual 0.0787617 0.28065   
## Number of obs: 1540, groups: pp, 96; Item, 90  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 5.43655 0.01329 409.122  
## Age\_PredY:L-H 0.03976 0.02056 1.934  
## Age\_PredO:L-H 0.04952 0.02035 2.434  
## Age\_Pred 0.12476 0.02581 4.834

# SINGLE FIXATION DURATION (SFD)

measure = "SFD"   
datafile$depvar = log(datafile[,measure])  
datafile$depvar = as.numeric(datafile$depvar)

## linear mixed effect model analysis for SFD

SFD\_model= lmer(depvar~Age\*Predictability + (1|pp) + (1+Predictability|Item),datafile)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 0.00334954 (tol = 0.002, component 1)

summary(SFD\_model)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: depvar ~ Age \* Predictability + (1 | pp) + (1 + Predictability |   
## Item)  
## Data: datafile  
##   
## REML criterion at convergence: 531.2  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.5016 -0.5973 0.0030 0.5791 3.6629   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr  
## pp (Intercept) 0.010163 0.10081   
## Item (Intercept) 0.001526 0.03907   
## Predictability2-1 0.001466 0.03829 0.31  
## Residual 0.076992 0.27747   
## Number of obs: 1395, groups: pp, 96; Item, 90  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 5.43411 0.01374 395.460  
## Age2-1 0.12198 0.02623 4.651  
## Predictability2-1 0.02848 0.01567 1.817  
## Age2-1:Predictability2-1 -0.01120 0.03026 -0.370  
##   
## Correlation of Fixed Effects:  
## (Intr) Age2-1 Prd2-1  
## Age2-1 0.167   
## Prdctblt2-1 0.024 0.005   
## Ag2-1:Pr2-1 0.005 0.001 0.055   
## optimizer (nloptwrap) convergence code: 0 (OK)  
## Model failed to converge with max|grad| = 0.00334954 (tol = 0.002, component 1)

confint(SFD\_model, method="Wald")

## 2.5 % 97.5 %  
## .sig01 NA NA  
## .sig02 NA NA  
## .sig03 NA NA  
## .sig04 NA NA  
## .sigma NA NA  
## (Intercept) 5.407173792 5.46103842  
## Age2-1 0.070577329 0.17338200  
## Predictability2-1 -0.002234261 0.05919447  
## Age2-1:Predictability2-1 -0.070516580 0.04811908

## simple effect analysis for SFD

contrast.matrix <- matrix(c(  
 -1, 1, 0, 0,   
 0, 0, -1, 1), 4, 2,  
 dimnames=list(c("YH","YL","OH","OL"),   
 c("Y:L-H", "O:L-H")))  
inv.contrast.matrix <- matrix(t(ginv(contrast.matrix)), 4, 2,  
 dimnames=list(c("YH","YL","OH","OL"),   
 c("Y:L-H", "O:L-H")))  
(contrasts(datafile$Age\_Pred) <- fractions(inv.contrast.matrix))

## Y:L-H O:L-H  
## YH -1/2 0   
## YL 1/2 0   
## OH 0 -1/2   
## OL 0 1/2

depvar.lmer1 = lmer(depvar ~ Age\_Pred + (1|pp)+(1|Item), datafile)  
summary(depvar.lmer1, corr = FALSE)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: depvar ~ Age\_Pred + (1 | pp) + (1 | Item)  
## Data: datafile  
##   
## REML criterion at convergence: 531.5  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.5093 -0.5945 0.0080 0.5802 3.6755   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## pp (Intercept) 0.010231 0.10115   
## Item (Intercept) 0.001522 0.03901   
## Residual 0.077336 0.27809   
## Number of obs: 1395, groups: pp, 96; Item, 90  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 5.43407 0.01377 394.614  
## Age\_PredY:L-H 0.03399 0.02081 1.633  
## Age\_PredO:L-H 0.02306 0.02203 1.046  
## Age\_Pred 0.12216 0.02629 4.646

# GAZE DURATION (GD)

measure = "GD"   
datafile$depvar = log(datafile[,measure])  
datafile$depvar = as.numeric(datafile$depvar)

## linear mixed effect model analysis for GD

GD\_model= lmer(depvar~Age\*Predictability + (1|pp) + (1|Item),datafile)  
summary(GD\_model)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: depvar ~ Age \* Predictability + (1 | pp) + (1 | Item)  
## Data: datafile  
##   
## REML criterion at convergence: 1173.6  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.2372 -0.6121 -0.0618 0.5244 4.1814   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## pp (Intercept) 0.0113786 0.1067   
## Item (Intercept) 0.0006151 0.0248   
## Residual 0.1164792 0.3413   
## Number of obs: 1540, groups: pp, 96; Item, 90  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 5.49645 0.01461 376.218  
## Age2-1 0.19728 0.02875 6.863  
## Predictability2-1 0.07198 0.01756 4.100  
## Age2-1:Predictability2-1 0.04837 0.03511 1.378  
##   
## Correlation of Fixed Effects:  
## (Intr) Age2-1 Prd2-1  
## Age2-1 0.143   
## Prdctblt2-1 -0.015 -0.004   
## Ag2-1:Pr2-1 -0.004 -0.016 -0.009

confint(GD\_model, method="Wald")

## 2.5 % 97.5 %  
## .sig01 NA NA  
## .sig02 NA NA  
## .sigma NA NA  
## (Intercept) 5.46781547 5.5250847  
## Age2-1 0.14094334 0.2536230  
## Predictability2-1 0.03757308 0.1063912  
## Age2-1:Predictability2-1 -0.02043862 0.1171782

## simple effect analysis for GD

contrast.matrix <- matrix(c(  
 -1, 1, 0, 0,   
 0, 0, -1, 1), 4, 2,  
 dimnames=list(c("YH","YL","OH","OL"),   
 c("Y:L-H", "O:L-H")))  
inv.contrast.matrix <- matrix(t(ginv(contrast.matrix)), 4, 2,  
 dimnames=list(c("YH","YL","OH","OL"),   
 c("Y:L-H", "O:L-H")))  
(contrasts(datafile$Age\_Pred) <- fractions(inv.contrast.matrix))

## Y:L-H O:L-H  
## YH -1/2 0   
## YL 1/2 0   
## OH 0 -1/2   
## OL 0 1/2

depvar.lmer1 = lmer(depvar ~ Age\_Pred + (1|pp)+(1|Item), datafile)  
summary(depvar.lmer1, corr = FALSE)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: depvar ~ Age\_Pred + (1 | pp) + (1 | Item)  
## Data: datafile  
##   
## REML criterion at convergence: 1173.6  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.2372 -0.6121 -0.0618 0.5244 4.1814   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## pp (Intercept) 0.0113786 0.1067   
## Item (Intercept) 0.0006151 0.0248   
## Residual 0.1164792 0.3413   
## Number of obs: 1540, groups: pp, 96; Item, 90  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 5.49645 0.01461 376.218  
## Age\_PredY:L-H 0.04780 0.02493 1.917  
## Age\_PredO:L-H 0.09617 0.02472 3.891  
## Age\_Pred 0.19728 0.02875 6.863

# TOTAL READING TIME (TRT)

measure = "TRT"   
datafile$depvar = log(datafile[,measure])  
datafile$depvar = as.numeric(datafile$depvar)

## linear mixed effect model analysis for TRT

TRT\_model= lmer(depvar~Age\*Predictability + (1|pp) + (1|Item),datafile)  
summary(TRT\_model)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: depvar ~ Age \* Predictability + (1 | pp) + (1 | Item)  
## Data: datafile  
##   
## REML criterion at convergence: 2598.1  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.1349 -0.6770 -0.1047 0.6438 3.6878   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## pp (Intercept) 0.024913 0.15784   
## Item (Intercept) 0.008718 0.09337   
## Residual 0.231760 0.48142   
## Number of obs: 1767, groups: pp, 96; Item, 90  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 5.70825 0.02265 252.044  
## Age2-1 0.37914 0.04080 9.294  
## Predictability2-1 0.09538 0.02319 4.114  
## Age2-1:Predictability2-1 0.12433 0.04635 2.682  
##   
## Correlation of Fixed Effects:  
## (Intr) Age2-1 Prd2-1  
## Age2-1 0.163   
## Prdctblt2-1 -0.015 -0.001   
## Ag2-1:Pr2-1 0.000 -0.017 0.053

confint(TRT\_model, method="Wald")

## 2.5 % 97.5 %  
## .sig01 NA NA  
## .sig02 NA NA  
## .sigma NA NA  
## (Intercept) 5.66386583 5.7526438  
## Age2-1 0.29918531 0.4590989  
## Predictability2-1 0.04993662 0.1408216  
## Age2-1:Predictability2-1 0.03348459 0.2151666

## simple effect analysis for TRT

contrast.matrix <- matrix(c(  
 -1, 1, 0, 0,   
 0, 0, -1, 1), 4, 2,  
 dimnames=list(c("YH","YL","OH","OL"),   
 c("Y:L-H", "O:L-H")))  
inv.contrast.matrix <- matrix(t(ginv(contrast.matrix)), 4, 2,  
 dimnames=list(c("YH","YL","OH","OL"),   
 c("Y:L-H", "O:L-H")))  
(contrasts(datafile$Age\_Pred) <- fractions(inv.contrast.matrix))

## Y:L-H O:L-H  
## YH -1/2 0   
## YL 1/2 0   
## OH 0 -1/2   
## OL 0 1/2

depvar.lmer1 = lmer(depvar ~ Age\_Pred + (1|pp)+(1|Item), datafile)  
summary(depvar.lmer1, corr = FALSE)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: depvar ~ Age\_Pred + (1 | pp) + (1 | Item)  
## Data: datafile  
##   
## REML criterion at convergence: 2598.1  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.1349 -0.6770 -0.1047 0.6438 3.6878   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## pp (Intercept) 0.024913 0.15784   
## Item (Intercept) 0.008718 0.09337   
## Residual 0.231760 0.48142   
## Number of obs: 1767, groups: pp, 96; Item, 90  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 5.70825 0.02265 252.044  
## Age\_PredY:L-H 0.03322 0.03190 1.041  
## Age\_PredO:L-H 0.15754 0.03364 4.683  
## Age\_Pred 0.37914 0.04080 9.294