UROP Final Report

Emi + Sophia

2023-05-06

Abstract

Cognitive Load Theory (Sweller, 2011) suggests that worked examples, mathematical problems with written out solutions, can increase students' available working memory resources - which are a limited set of cognitive attention resources that allow students to learn mathematics concepts and solve word problems. Worked examples may offload attention to written word problem solutions rather than requiring learners to hold them in mind, creating more efficient learning. At the same time, many students feel mathematical anxiety when faced with mathematics learning and problem solving, which can generate worries that engage the same limited set of working memory resources, reducing available resources for new learning and problem solving. To test whether worked examples could mitigate the deleterious effects of mathematical anxiety on performance, a pre-post design experiment was conducted on 280 fifth-graders from ten schools in Orange County and Chicago to see whether worked examples could reduce mathematical anxiety in students learning ratios - a challenging mathematical concept. The study also examined other potential relationships with mathematical anxiety such as mind wandering and situational interest. A multiple linear regression model was used to assess the relationships between these key features. Findings indicate that for those who were given worked examples during instruction, their mathematical anxiety was mitigated which allowed them to perform more accurately in the delayed assessment. Mathematical anxiety was a significant predictor of mind wandering, which is a factor in predicting students' performance as well.

Study Summary

- 280 fifth-graders from 10 schools, 8 in Orange County, California and 2 in Chicago, Illinois

- **Pre-post design experiment:** students were given a 20-minute video lesson on ratios on day 1

  - Experimental group: worked examples present
  - Control group: no worked examples provided

- **Post-study** (3 days after day 1): Students completed math anxiety questionnaire (TMA) and post-test assessing their ratio knowledge

Load Libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.4.0     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1


## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'


## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(ggpubr)
```

Load Data

```
dfr <- readxl::read_xlsx("../data/KidsCOVID_Winter21_raw data.xlsx")
```

```
## New names:
## * 'Subject ID #' -> 'Subject ID #...8'
## * 'Subject ID #' -> 'Subject ID #...9'
## * 'StartDate' -> 'StartDate...11'
## * 'EndDate' -> 'EndDate...12'
## * 'Progress' -> 'Progress...13'
## * 'Duration (in seconds)' -> 'Duration (in seconds)...17'
## * 'Finished' -> 'Finished...18'
## * 'RecordedDate' -> 'RecordedDate...19'
## * 'Browser_info_Browser' -> 'Browser_info_Browser...20'
## * 'Browser_info_Version' -> 'Browser_info_Version...21'
## * 'Browser_info_Operating System' -> 'Browser_info_Operating System...22'
## * 'Browser_info_Resolution' -> 'Browser_info_Resolution...23'
## * 'Misc_0' -> 'Misc_0...36'
## * 'Misc_1' -> 'Misc_1...100'
## * 'Misc_2' -> 'Misc_2...145'
## * 'MW_control_1' -> 'MW_control_1...210'
## * 'MW_control_2' -> 'MW_control_2...211'
## * 'MW_control_3' -> 'MW_control_3...212'
## * 'MW_control_4' -> 'MW_control_4...213'
## * 'MW_control_5' -> 'MW_control_5...214'
## * '' -> '...219'
## * 'StartDate' -> 'StartDate...221'
## * 'EndDate' -> 'EndDate...222'
## * 'Progress' -> 'Progress...223'
```

```
## * 'Duration (in seconds)' -> 'Duration (in seconds)...224'
## * 'Finished' -> 'Finished...225'
## * 'RecordedDate' -> 'RecordedDate...226'
## * 'Browser_info_Browser' -> 'Browser_info_Browser...227'
## * 'Browser_info_Version' -> 'Browser_info_Version...228'
## * 'Browser_info_Operating System' -> 'Browser_info_Operating System...229'
## * 'Browser_info_Resolution' -> 'Browser_info_Resolution...230'
## * 'Misc_0' -> 'Misc_0...232'
## * 'Misc_1' -> 'Misc_1...239'
## * 'Misc_2' -> 'Misc_2...290'
## * 'MW_control_1' -> 'MW_control_1...295'
## * 'MW_control_2' -> 'MW_control_2...296'
## * 'MW_control_3' -> 'MW_control_3...297'
## * 'MW_control_4' -> 'MW_control_4...298'
## * 'MW_control_5' -> 'MW_control_5...299'
```

```r
# Number of rows in dfr
nrow(dfr)
```

```
## [1] 303
```

```r
# Number of columns in dfr
ncol(dfr)
```
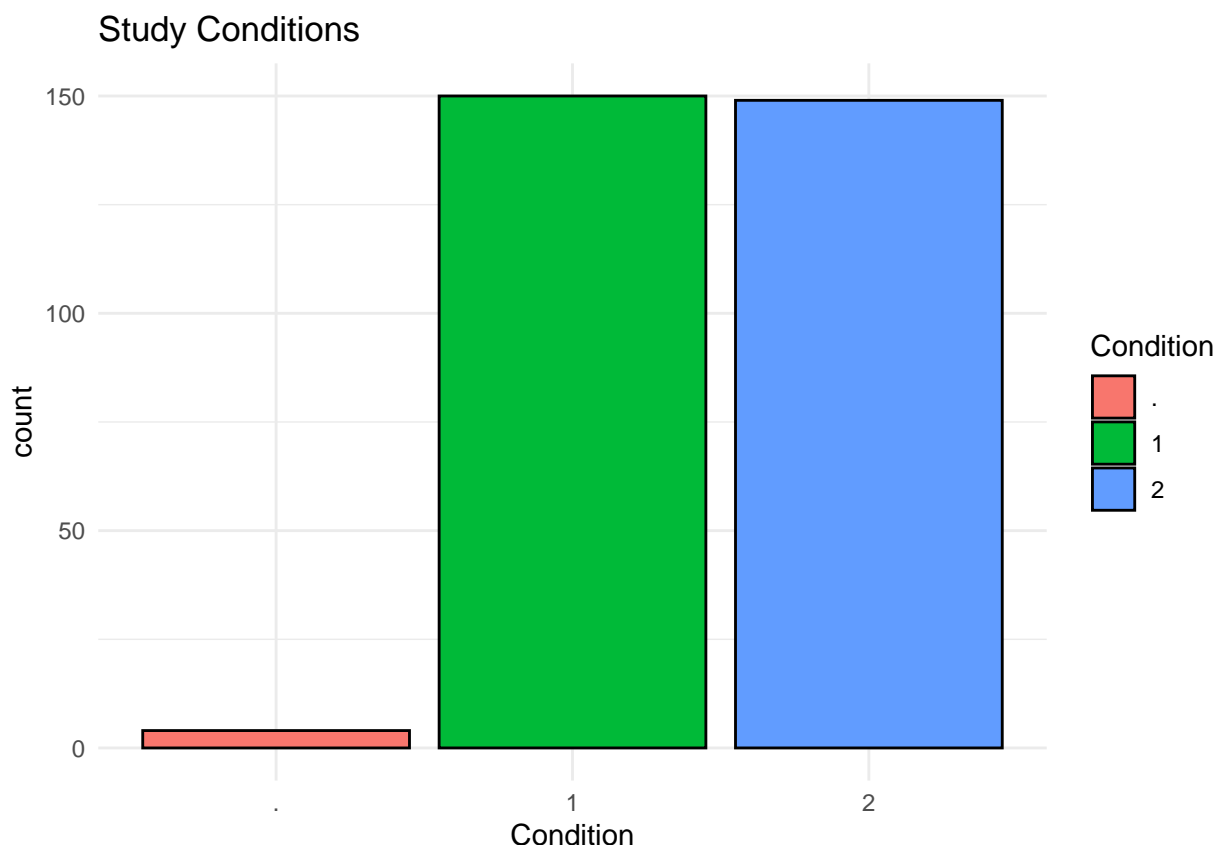
```
## [1] 409
```

The raw data contains 299 students' data and 410 vairables

Data Cleaning

We first noticed that some students from the data were not assigned to any group (NA):

```r
dfr %>% ggplot(aes(x = Condition, fill = Condition)) +
  geom_bar(color = "black") + theme_minimal() +
  labs(title = "Study Conditions")
```

Study Conditions

We want to look at *no worked example* and *worked example* group, so we removed all the data without condition. The raw data assigned `Condition = 1` is for students in *worked example* group, and `Condition = 2` for students in *no worked example* group. For clear readability, we assigned 0 to all students in *no worked example* group:

```
# removing NA condition data
dfr <-dfr %>% filter(Condition == 1 | Condition == 2)
# assigning 0 to no worked example group data
dfr$Condition[dfr$Condition == 2] = 0
```

For our UROP project, we are interested in comparing the effects of math anxiety (`TMA_avg`) on students' accuracy scores (`Del_OverallAcc`) between the worked example and no worked example groups. Furthermore, we are interested in observing the impacts of other factors from data such as mind wandering scores (`mw_avg`) and situational interest (`SI_avg`).

```
# Get the average of mw_1 and mw_2
dfr <- dfr %>% mutate(mw_avg = rowMeans(dfr[346:347], na.rm = TRUE))
# Select columns and remove any row with missing data
df <- dfr %>% select(Del_OverallAcc, TMA_avg, Condition, mw_avg, SI_avg) %>%
  drop_na()
glimpse(df)
```

```
## Rows: 229
## Columns: 5
## $ Del_OverallAcc <dbl> 0.3666667, 0.9000000, 0.5000000, 0.6000000, 0.1666667, ~
## $ TMA_avg        <dbl> 1.833333, 2.000000, 2.166667, 1.666667, 2.000000, 1.666~
## $ Condition      <chr> "1", "1", "1", "1", "0", "1", "1", "0", "0", "0", "1", ~
```

```
## $ mw_avg          <dbl> 1.9, 1.9, 2.3, 2.0, 1.5, 1.2, 2.1, 1.2, 2.6, 1.1, 1.2, ~
## $ SI_avg          <dbl> 4.500000, 3.000000, 3.250000, 4.500000, 4.750000, 4.583~
```

The new data contains 229 students' data and 5 variables

Visualizations

Math Anxiety Distributions

```r
# New facet label names for supp variable
cond.labs <- c("No Worked EX", "Worked EX")
names(cond.labs) <- c(0, 1)
# Histogram of tma_avg
df %>%
  ggplot(aes(x = TMA_avg, fill = Condition)) +
  geom_histogram(color = "black", bins = 20) +
  theme_minimal()+
  scale_fill_manual(values = c("#ef476f","#1b9aaa"), labels = c("No Worked EX", "Worked EX")) +
  facet_grid(.~Condition, labeller = labeller(Condition = cond.labs))  +
  labs(title = "Math Anxiety Distributions", x = "TMA Average")
```
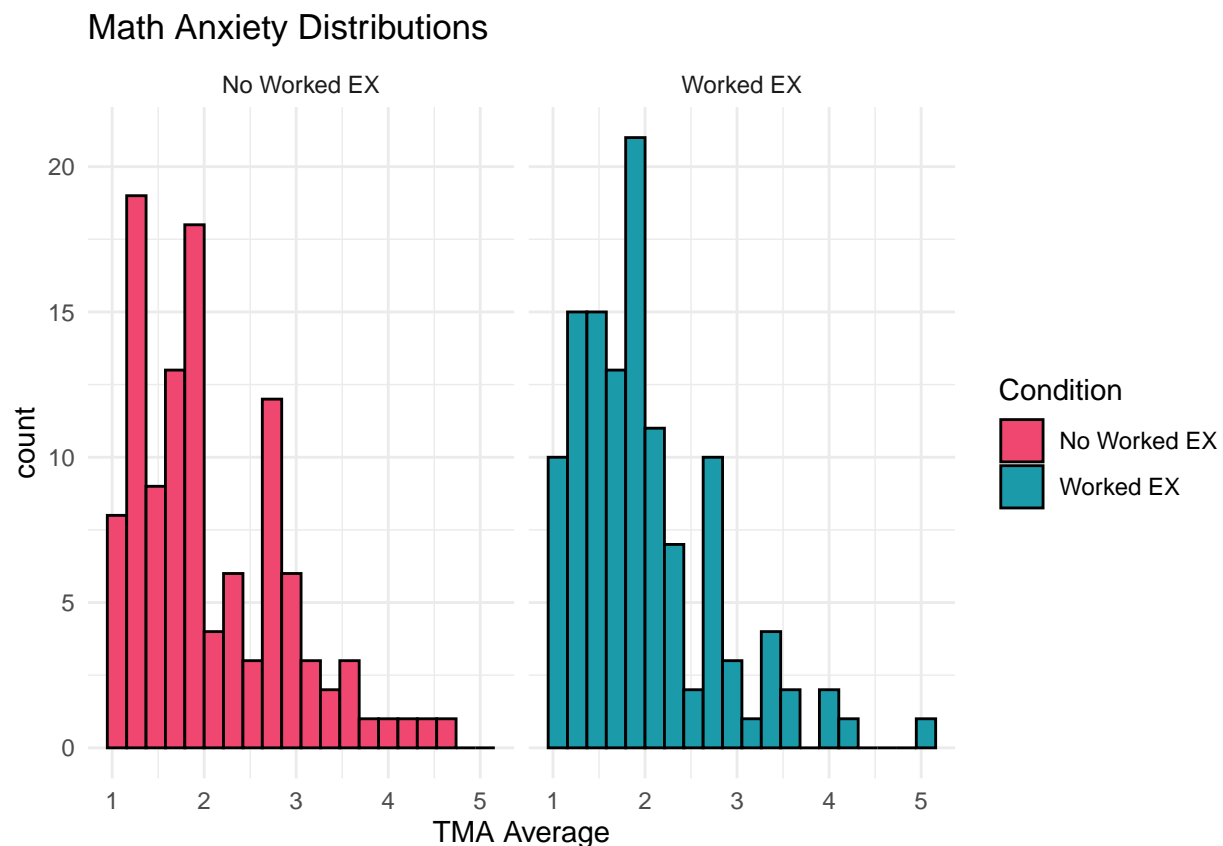

For both groups, the distributions seemed to be right-skewed, indicating that the average math anxiety was low for both groups:

```
# table summary of math anxiety
df %>%
  group_by(Condition) %>%
  summarize("Average TMA" = mean(TMA_avg), "Standard deviation" = sd(TMA_avg))
```

```
## # A tibble: 2 x 3
##   Condition 'Average TMA' 'Standard deviation'
##   <chr>             <dbl>                <dbl>
## 1 0                  2.09                0.832
## 2 1                  1.99                0.762
```
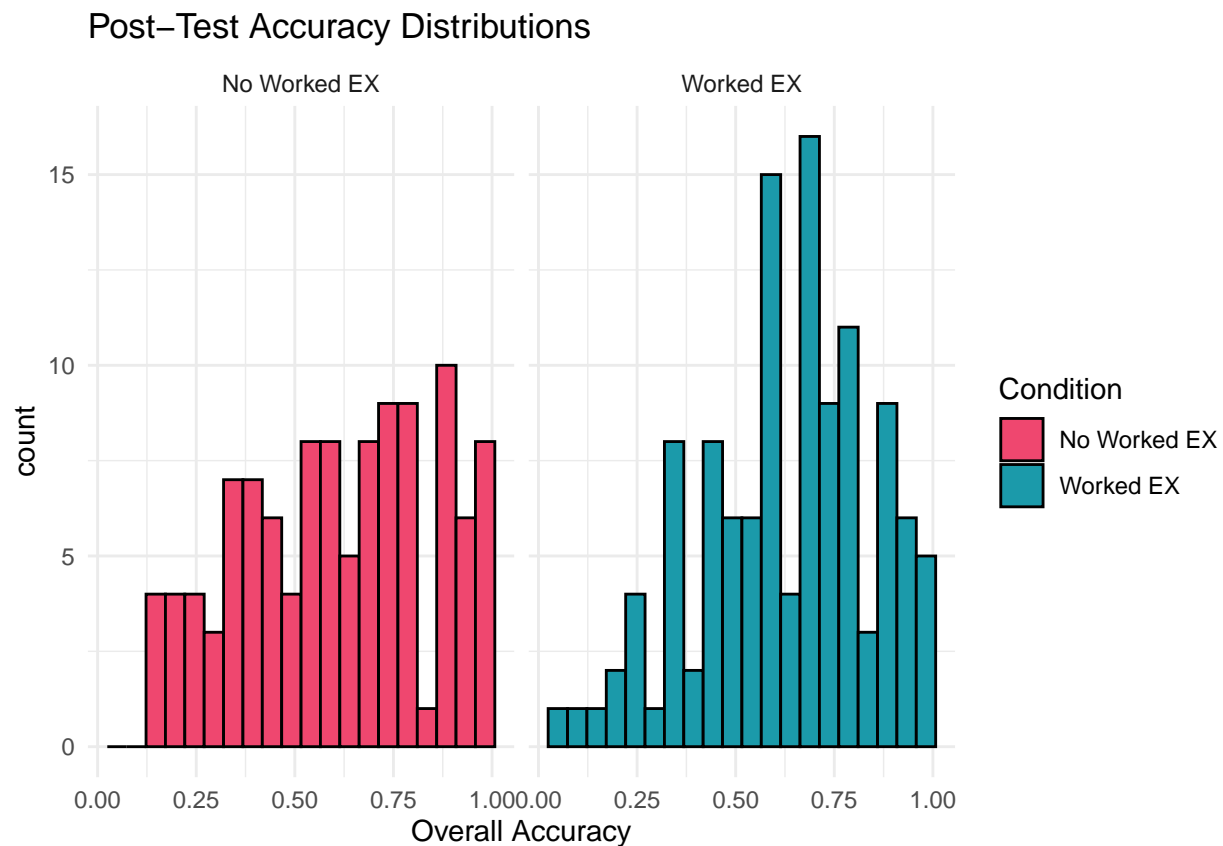
```
# Math anxiety average bar plot
df %>%
  group_by(Condition) %>%
  summarize(tma_avg = mean(TMA_avg))  %>%
  ggplot(aes(x = Condition, y = tma_avg, fill = Condition)) +
  geom_bar(position="dodge", stat="identity") + theme_bw() +
  scale_x_discrete(breaks=c(0,1),
        labels=c("No Worked EX","Worked EX")) +
  scale_fill_manual(values = c("#ef476f","#1b9aaa"), labels =
                           c("No Worked EX", "Worked EX")) +
  labs(x = "Condition", y = "TMA", title = "Average Math Anxiety Scores")+ ylim(0,5)
```
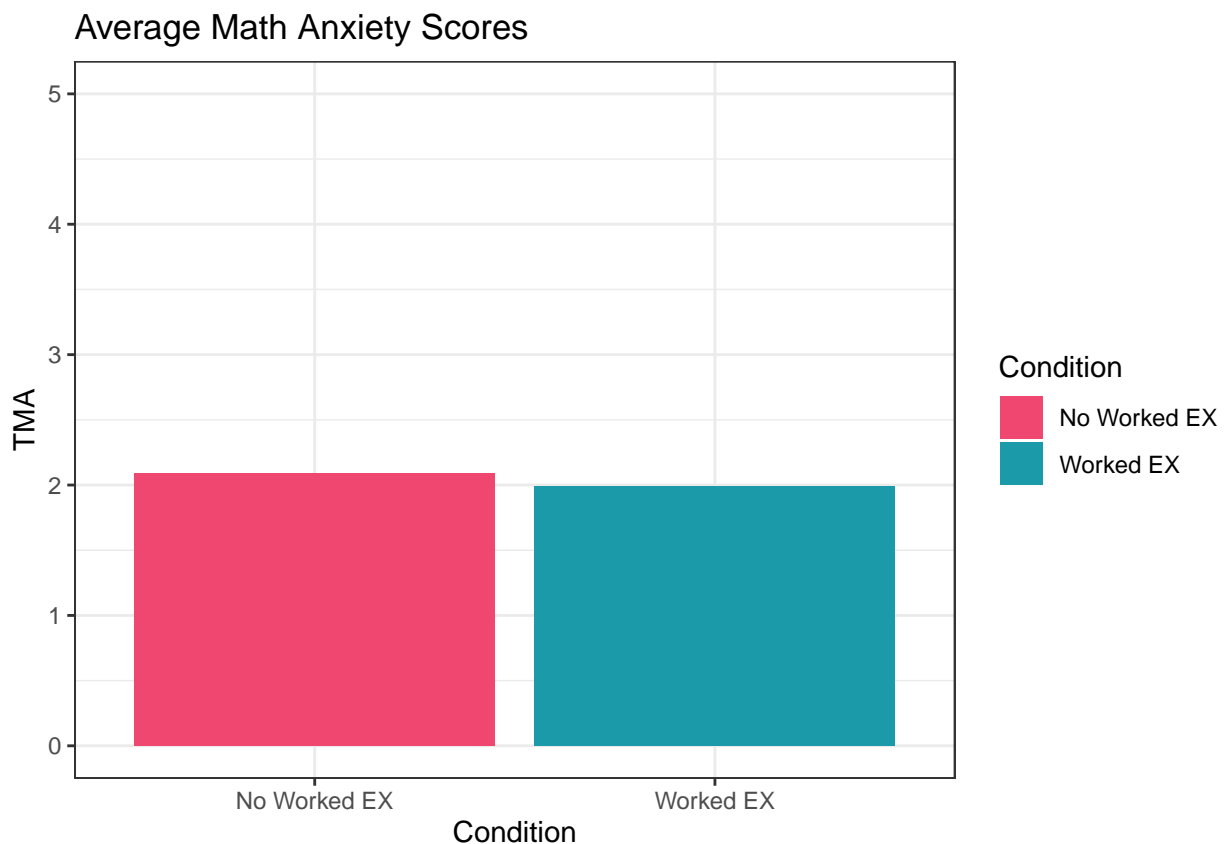
Post-test accuracy score distributions

```r
# histogram of post-assessment accuracy score
df %>%
  ggplot(aes(x = Del_OverallAcc, fill = Condition)) +
  geom_histogram(color = "black", bins = 20) +
  theme_minimal()+
  scale_fill_manual(values = c("#ef476f","#1b9aaa"), labels = c("No Worked EX", "Worked EX")) +
  facet_grid(.~Condition, labeller = labeller(Condition = cond.labs))  +
  labs(title = "Post-Test Accuracy Distributions", x = "Overall Accuracy")
```


The averages of post-test accuracy seem to be close for both groups:

```r
# table summary of accuracy
df %>%
  group_by(Condition) %>%
  summarize("Average Accuracy" = mean(Del_OverallAcc), "Standard deviation" = sd(Del_OverallAcc))
```

```
## # A tibble: 2 x 3
##   Condition 'Average Accuracy' 'Standard deviation'
##   <chr>                  <dbl>                <dbl>
## 1 0                      0.610                0.239
## 2 1                      0.627                0.214
```

```
# test accuracy average bar plot
df %>%
  group_by(Condition) %>%
  summarize(acc_avg = mean(Del_OverallAcc))  %>%
  ggplot(aes(x = Condition, y = acc_avg, fill = Condition)) +
  geom_bar(position="dodge", stat="identity") + theme_bw() +
  scale_x_discrete(breaks=c(0,1),
        labels=c("No Worked EX","Worked EX")) +
  scale_fill_manual(values = c("#ef476f","#1b9aaa"), labels =
                                c("No Worked EX", "Worked EX")) +
  labs(x = "Condition", y = "TMA", title = "Post-Test Overall Accuracy Scores") + ylim(0,1)
```
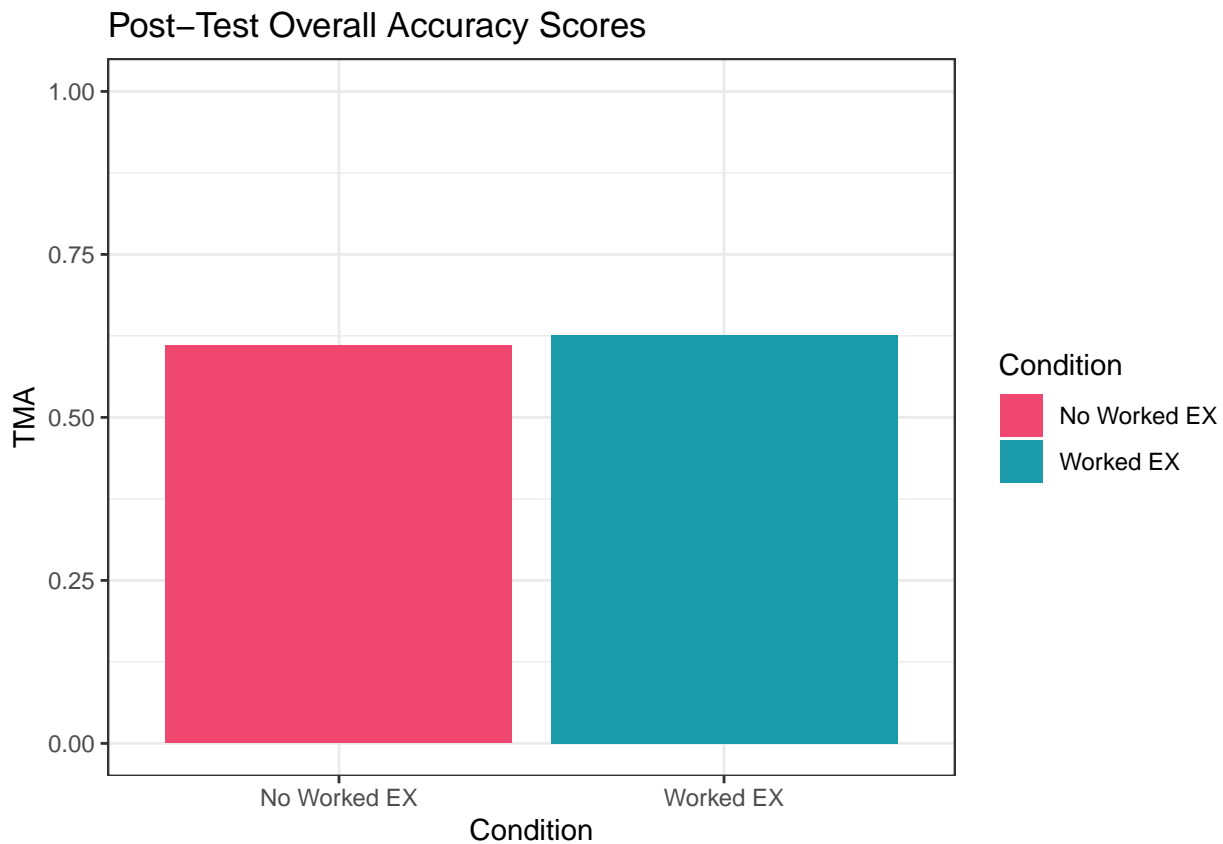

Relationship between math anxiety and overall accuracy

```
# Scatterplot: tma vs accuracy
df %>%
  ggplot(aes(x = TMA_avg, y = Del_OverallAcc ,color  = Condition)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_bw() +
  scale_color_manual(values = c("#ef476f","#1b9aaa"), labels =
                                c("No Worked EX", "Worked EX")) +
  facet_grid(.~Condition, labeller = labeller(Condition = cond.labs))  +
  labs(title = "Effect of Math Anxiety on Overall Accuracy",
                        x = "TMA Average", y = "Overall Accuracy")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Effect of Math Anxiety on Overall Accuracy
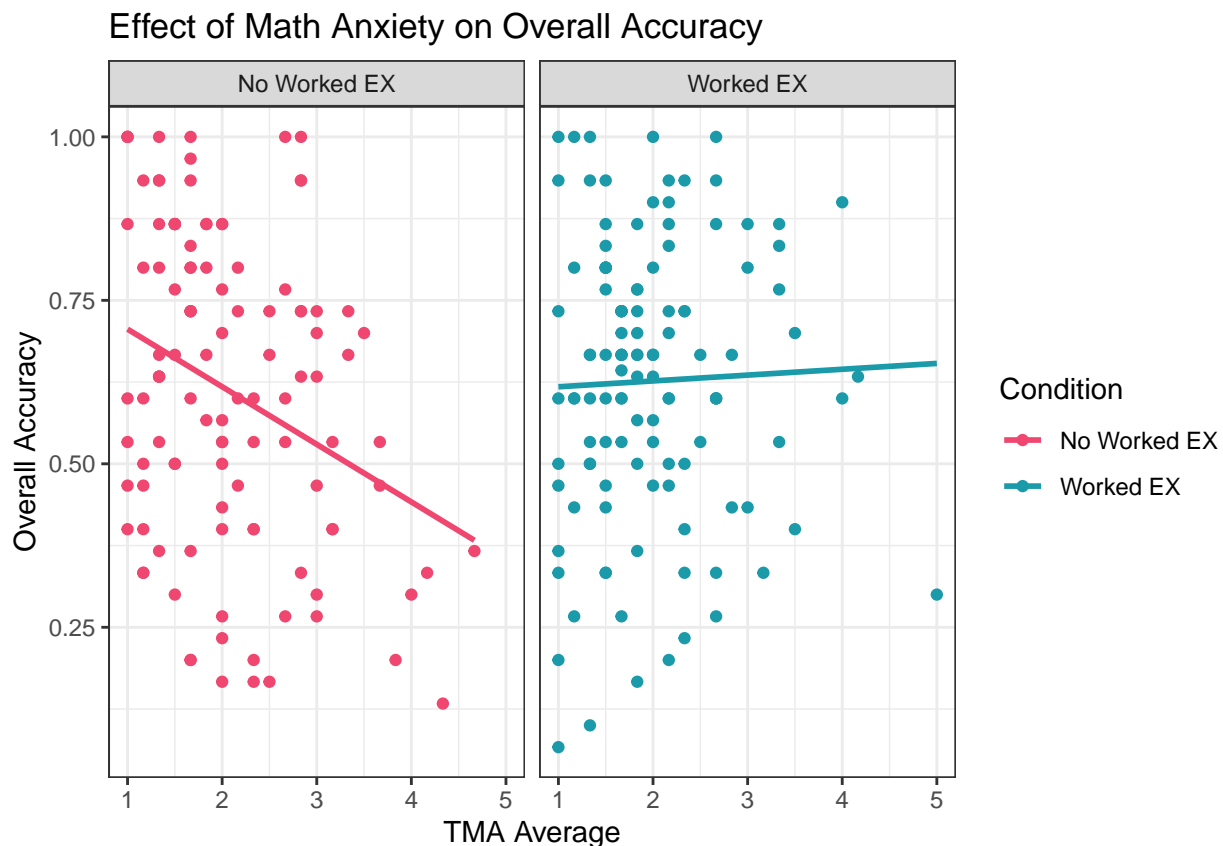
Based on the regression lines, we see a decrease in students' accuracy score as their math anxiety scores increase in the no worked example group (left), and we don't seem much of a change in students' accuracy scores in the worked example group (right).

Relationship between mind wandering / situational interest and math anxiety

```
# MW vs tma
mw_acc1 <- dfr %>%
        ggplot(aes(x = mw_avg, y = TMA_avg, color = Condition)) +
        geom_point() + geom_smooth(method = "lm", se = FALSE) +
        theme_bw() +
        scale_color_manual(values = c("#ef476f","#1b9aaa"), labels =
                                c("No Worked EX", "Worked EX")) +
        scale_fill_brewer(labels = c("No Worked EX", "Worked EX")) +
        labs(x = "Mind Wandering", y = "TMA")
#SI vs tma
si_acc1 <- dfr %>%
  ggplot(aes(x = SI_avg, y = TMA_avg, color = Condition)) +
  geom_point() + geom_smooth(method = "lm", se = FALSE) +
  theme_bw() +
  scale_color_manual(values = c("#ef476f","#1b9aaa"), labels = c("No Worked EX", "Worked EX")) +
```

```
  labs(x = "Situational Interest", y = "TMA")
# Side-by-side plot mw and situational interests with understanding
ggarrange(mw_acc1, si_acc1, common.legend = TRUE, legend="bottom", ncol = 2, nrow = 1)
```

'geom_smooth()' using formula = 'y ~ x'

Warning: Removed 59 rows containing non-finite values ('stat_smooth()').

Warning: Removed 59 rows containing missing values ('geom_point()').
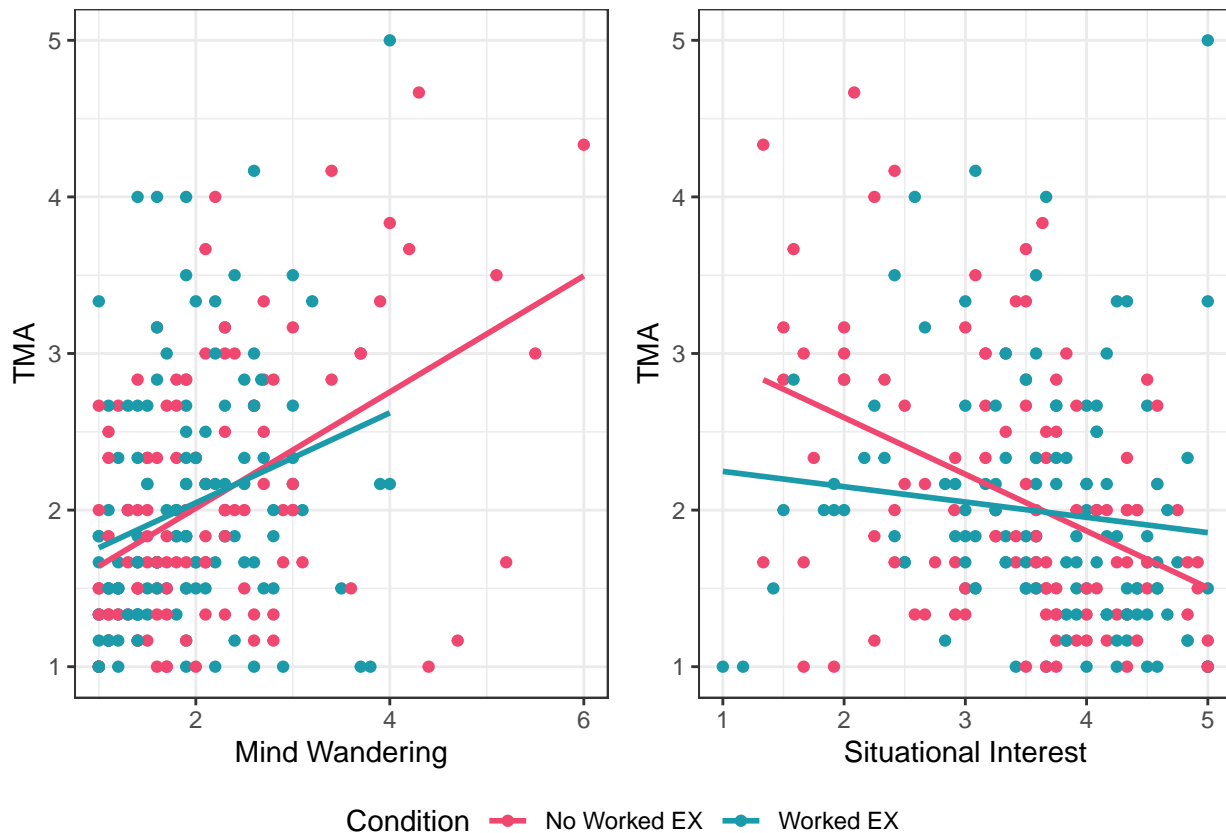
'geom_smooth()' using formula = 'y ~ x'

Warning: Removed 59 rows containing non-finite values ('stat_smooth()').
Removed 59 rows containing missing values ('geom_point()').

'geom_smooth()' using formula = 'y ~ x'

Warning: Removed 70 rows containing non-finite values ('stat_smooth()').

Warning: Removed 70 rows containing missing values ('geom_point()').

Based on the regression lines, we see a positive relationship between mind wandering and math anxiety (left) and a negative relationship between situational interest and math anxiety (right).

Modeling

Predicting student's achievements (delayed accuracy scores)

Dependent variable: Post-study accuracy scores (`Del_OverallAcc`)

Independent variables: Math anxiety (`TMA_avg`), condition (`Condition`), interaction term between math anxiety and condition (`Condition * TMA_avg`), mind wandering avg (`mw_avg`), situational interest (`SI_avg`)

```
acc_fit <- df %>%
  lm(formula = Del_OverallAcc ~ TMA_avg + Condition +  Condition * TMA_avg + mw_avg + SI_avg)
summary(acc_fit)
```

```
##
## Call:
## lm(formula = Del_OverallAcc ~ TMA_avg + Condition + Condition *
##     TMA_avg + mw_avg + SI_avg, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.53186 -0.16809  0.01063  0.17062  0.43554
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.851342   0.096854   8.790 4.08e-16 ***
## TMA_avg           -0.067673   0.028332  -2.389   0.0177 *
## Condition1        -0.171011   0.081288  -2.104   0.0365 *
## mw_avg            -0.037099   0.018370  -2.020   0.0446 *
## SI_avg            -0.005478   0.017012  -0.322   0.7477
## TMA_avg:Condition1 0.086362   0.037650   2.294   0.0227 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2209 on 223 degrees of freedom
## Multiple R-squared:  0.06932,    Adjusted R-squared:  0.04845
## F-statistic: 3.322 on 5 and 223 DF,  p-value: 0.006491
```

Significant factors for predicting students' accuracy:

- `TMA_avg` (β = -0.067673, p = 0.0177): Increased in math anxiety caused decrease in students' accuracy

- `Condition1` (β = -0.171011, p = 0.0365): Worked example group had decrease in accuracy by 0.187 percent

- `mw_avg` (β = -0.037099, p = 00.0446): Mind wandering had negative effect on student's accuracy score. (decrease by 0.037099)

- `TMA_avg:Condition1` (β = 0.086362, p = 0.0227): The math anxiety scores of students in worked example group had change in accuracy by 0.100 percent -> There is a statistical evidence that indicates that effects of math anxiety on accuracy scores differ among the study groups

Model for worked example group:

$$DelOverallAcc = 0.680331 + 0.018689 * TMAavg - 0.037099 * MWavg - 0.005468 * SIavg$$

Model for no-worked example group:

$$DelOverallAcc = 0.851342 - 0.067673 * TMAavg - 0.037099 * MWavg - 0.005468 * SIavg$$

Math anxiety has different impacts on students' achievements among the study groups. In worked example group, there is a positive effect on students' accuracy scores by math anxiety and in no-worked example group, there is a negative effect on students' accuracy scores by math anxiety.

Predicting student's math anxiety

Dependent variable: Mind wandering avg (`mw_avg`)

Independent variables: Math anxiety (`TMA_avg`), condition (`Condition`), interaction term between math anxiety and condition (`Condition * TMA_avg`), situational interest (`SI_avg`)

```
mw_fit <- dfr %>%
  lm(formula = mw_avg ~ TMA_avg + Condition + TMA_avg * Condition + SI_avg)
summary(mw_fit)
```

```
##
## Call:
## lm(formula = mw_avg ~ TMA_avg + Condition + TMA_avg * Condition +
##     SI_avg, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4683 -0.5322 -0.1488  0.3225  3.1675
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1.55264    0.33665   4.612 6.71e-06 ***
## TMA_avg            0.54954    0.09629   5.707 3.61e-08 ***
## Condition1         0.38260    0.29455   1.299   0.1953
## SI_avg            -0.14973    0.06106  -2.452   0.0150 *
## TMA_avg:Condition1 -0.28699    0.13559  -2.117   0.0354 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8034 on 224 degrees of freedom
##   (70 observations deleted due to missingness)
## Multiple R-squared:  0.2282, Adjusted R-squared:  0.2144
## F-statistic: 16.55 on 4 and 224 DF,  p-value: 6.72e-12
```

- `TMA_avg` ($\beta = 0.54954$, p = 3.61e-08): Increase in math anxiety cause increase in mind wandering

- `TMA_avg:Condition1`(β = -0.28699, p = 0.03540): In worked example group, increase in math anxiety cause decrease in mind wandering -> There is a statistical evidence indicating that effects of math anxiety on mind wandering differ among the study groups

- `SI_avg` (β = -0.14973, p = 0.0150): Increase in situational interest caused decrease in students' mind wandering

Model for worked example group:

$$MWavg = 1.93524 + 0.26255 * TMA - 0.14973 * SIavg$$

Model for no-worked example group:

$$MWavg = 1.55264 + 0.54954 * TMA - 0.14973 * SIavg$$

Math anxiety has different impacts on mind wandering among study groups. Although both groups showed positive effects of math anxiety on mind wandeirng, in the worked example group, the effect of math anxiety on mind wandering is smaller than the no worked example group.