COS-D419 Factor Analysis and Structural Equation Models 2023, Assignment 4

Rong Guang

2023-02-11

Contents

1 SEM & teacher burnout

1.1 Exercise 4.1

Draw the graph of the initial, full structural equation model. Make sure that you have included all the specified paths.

Estimate the initial model using the robust MLM estimator *(robust variant of the ML estimator, to be precise!)* and present a brief summary of the model fit.

2 Preparation

2.1 Read in the data set

Start by downloading the data file from Moodle to Project folder.

```
library(tidyverse)
library(readr)
library(here)

#This week's file name
latest.name <- "ALLSEC.CSV"

#read in the data
mbi <- read_csv(file.path(here(),
                'data',
                latest.name))
```

2.2 Write functions

To control length of reports, codes already shown in the previous homework were not showing in the current report. Yet they are available in .rmd report.

2.2.1 to check unique values

2.2.2 to generate CFA results with improved readability

2.2.3 to generate functions for improving aethetics of correlation matrix

2.2.4 to generate a function for histogram overlapping with density plot

2.2.5 to generate a function for violin overlapping with box plot

2.2.6 To generate a function describing continuous data set

2.2.7 Write a function to print a table with concerned parameters

```
#write a function for minus calculation
minus <- function(x,y) {x - y}

#write a function to print a table with concerned parameters
concern.table <- function(sem, nofpath, model){
options(scipen = 999)
#This is for structural path residual variance
##regression path estimates
sem.parameter <- parameterEstimates(sem, standardized=TRUE) |>  # obtain estimates
  filter(op == "~") |>   #select "is measured by" rows
  mutate(Parameter = paste0(rhs, "→", lhs)) |>
  select(Parameter,
         'B'=est, #estimates
```

```r
         'Beta'=std.all,#estimates standardized
         SE=se, #standard error
         Z=z, #z statistics
         'p-value'=pvalue #p value
         )

##round the p-value column
sem.parameter$`p-value` <- sem.parameter$`p-value` |>
  round(3)

##add a conditional logic to the p-value column that >0.05 cell shows in red
sem.parameter$`p-value` <- cell_spec(sem.parameter$`p-value`,
                                     color = ifelse(
                                       sem.parameter$`p-value` > 0.05,
                                       "red",
                                       "black")
                                     )

#This is for the residual variance of dependence variable
##obtain estimates
variance <- parameterEstimates(sem1, standardized=TRUE)  |>
  filter(op == "~~") #select "is correlated with" rows
##subset needed rows (variance row)
variance <- variance |>
  filter(rhs %in% c("F8SELF", "F9ELC", "F10EE", "F11DP", "F12PA"))

#variance[minus(sum(32,nofpath), 5-1):sum(32,nofpath),] #32 is the n of indicators;
                                             #12 is the number of factors;
                                             #5 is the new of row I plan to show

##select&rename columns
sem.tab.variance <- variance |> select(
                   Parameter=rhs, #right hand side column
                   'B'=est, #estimates
                   'Beta'=std.all, #standardized estimates
                   SE=se,#standard error
                   Z=z, #z statistics
                   'p-value'=pvalue #p value
                   )
##remove the row names
rownames(sem.tab.variance) <- NULL
##round the p-value column
sem.tab.variance$`B` <- sem.tab.variance$`B` |>
  round(3)
##add a conditional logic to the p-value column that >0.05 cell shows in red
sem.tab.variance$`B` <- cell_spec(sem.tab.variance$`B`,
                                     color = ifelse(
                                       sem.tab.variance$`B` < 0,
                                       "red",
                                       "black")
                                     )

#bind the two table
```

```r
concern.table <- rbind(sem.parameter, sem.tab.variance)
concern.table[1:nofpath, 2] <-
  as.character(round(as.numeric(concern.table[1:nofpath, 2]),3))
concern.table[sum(nofpath, 1):sum(nofpath, 5),6] <-
  as.character(round(
    as.numeric(
      concern.table[sum(nofpath,1):sum(nofpath, 5),6]
      ),
    3)
    )


#further aesthetics
concern.table |>
  select("Parameter*" = Parameter,
         'B†' = B, #estimates
         'Beta‡' = Beta,#estimates standardized
         SE, #standard error
         Z, #z statistics
         'p-value') |> #p value
  kable(digits = 3, #rounded to 3
        #format="latex", #Latex markdown
        booktabs=TRUE, #Latex booktabs
        linesep = "",
        align = "lrrrrr",
        caption=
          paste(
            "Residual variance of structural regression path and select factors for",
            model),
        escape = F) |> #caption
  kable_styling(latex_options = "striped") |> #gray every other row
  pack_rows("Strutural regression path",
            1,nofpath) |>
  pack_rows("Selected factors",
            sum(nofpath,1), sum(nofpath,5)) |>
  footnote(general = "Values highlighted in red should be taken note of",
           symbol = c("→ indicates regression path",
                      "Crude estimates",
                      "Standardized estimates"))
}
```

2.2.8 To generate a function for calculating 2 difference was defined.

```r
chi.diff <- function(sem1, sem2){ #2 augments input: the 1st is the nested model
                                  #the 2nd is comparison model (less restricted)
  measure0 <- fitMeasures(sem1,  #extract robust fit indices for the 1st model
                          c("df.scaled",
                            "chisq.scaling.factor",
                            "chisq.scaled")) |>
    as.vector()   #turn to vector to facilitate calculation
  measure1 <- fitMeasures(sem2,  #extract robust fit indices for the 2nd model
                          c("df.scaled",
```

```r
                              "chisq.scaling.factor",
                              "chisq.scaled")) |>
    as.vector()
  cd <-  # this is the formula from p 14, week 4 slide
    (measure0[1]*measure0[2]-measure1[1]*measure1[2])/(measure0[1]-measure1[1])
  TRd <- # this is the formula from p 15, week 4 slide
    (measure0[3]*measure0[2]-measure1[3]*measure1[2])/cd
  TRd <- TRd |> round(3) # round the value to 3 decimal places
}
```

2.2.9 To generate a function for ploting full sem diagram

```r
#plot the diagram
full.sem.diagram <- function(model, nofpath, fig.num, quotedmodel){
semPaths(semPlotModel(model),
             style = "lisrel",
             rotation = 2,
             sizeLat = 6,
             sizeLat2 = 5,
             sizeMan = 5,
             sizeMan2 = 2,
             residScale = 4,
             shapeMan = "rectangle",
             edge.color = c(rep("black",34),
                            rep("blue", minus(nofpath,1)),
                            "orange", # this is the new free parameter
                            rep("gray",32),
                            rep("red",5)),
         residuals = T,
         layout = m,
         nCharNodes=0,
         optimizeLatRes = T,
         exoVar = F)
#add title
title(main = list(paste("Figure",
                        fig.num,
                        "Modified model (",
                        quotedmodel,
                        ") of teacher burnout"),
                  cex = 1.5, font =1),
     outer = F, line = -1)
#add notes
title(sub =
        "Notes: Orange arrow is the regression path set free to estimate in the current model",
      line = 0, adj = 0.7)

}
```

2.3 Inspect the data

2.3.1 Data structure

Have a quick overview of the data structure

```
library(knitr)
library(broom)
dim(mbi);mbi %>% apply(2, function(x)class(x));
```

```
## [1] 1430   32
```

```
##    ROLEA1    ROLEA2    ROLEC1    ROLEC2     WORK1     WORK2    CCLIM1    CCLIM2
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##    CCLIM3    CCLIM4      DEC1      DEC2     SSUP1     SSUP2     PSUP1     PSUP2
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##     SELF1     SELF2     SELF3      ELC1      ELC2      ELC3      ELC4      ELC5
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##       EE1       EE2       EE3       DP1       DP2       PA1       PA2       PA3
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
```

The data set contains 22 numeric variables of 372 obs. Their values appear to follow a consistent pattern covering the integer from 1 to 7, except for Items 4, 7, 17 and 21, which did not include a value of 1.

2.3.2 Descriptive statistics of measured variables

```
library(finalfit);library(kableExtra);

descriptive(mbi) |>
  pack_rows(index =
              c("Factor 1*: Role Ambiguity \n(high score means negative)" = 2,
                "Factor 2*: Role conflict \n(high score means negative)" = 2,
                "Factor 3*: Work overload \n(high score means negative)" = 2,
                "Factor 4‡: classroom climate" = 4,
                "Factor 5*: Decision-making" = 2,
                "Factor 6*: Superior support" =  2,
                "Factor 7*: Peer support" = 2,
                "Factor 8‡: Self-esteem" = 3,
                "Factor 9‡: External locus of control" = 5,
                "Factor 10†: Emotional Exhaustion \n(high score means negative)" = 3,
                "Factor 11†: Depersonalization \n(high score means negative)" = 2,
                "Factor 12†: Personal Accomplishment" = 3)) |>
  footnote(general =
             "Indicators variables were formulated through item parcels.",
    symbol = c("These indicators are parcels from Teacher Stress Scale instrument",
               "These indicators are parcels from BMI instrument",
               "These parcels consist of items from single unidimensional scales")
           )
```

Table 1: Descriptive statistics for measurements

	n	n of NA	Central tendency		Dispersion tendency			
			Mean	Median	SD	Min	Max	Q1~Q3
Factor 1*: Role Ambiguity (high score means negative)								
ROLEA1	1430	0	2.4	2.3	0.9	1.0	6.0	1.7 ~ 3.0
ROLEA2	1430	0	2.1	2.0	1.0	1.0	6.0	1.5 ~ 2.5
Factor 2*: Role conflict (high score means negative)								
ROLEC1	1430	0	3.0	3.0	1.1	1.0	6.0	2.3 ~ 3.7
ROLEC2	1430	0	3.0	3.0	1.2	1.0	6.0	2.0 ~ 4.0
Factor 3*: Work overload (high score means negative)								
WORK1	1430	0	3.2	3.3	1.2	1.0	6.0	2.3 ~ 4.0
WORK2	1430	0	2.2	2.0	1.1	1.0	6.0	1.5 ~ 3.0
Factor 4‡: classroom climate								
CCLIM1	1430	0	3.0	3.0	0.5	1.0	4.0	2.7 ~ 3.3
CCLIM2	1430	0	2.7	2.7	0.6	1.0	4.0	2.3 ~ 3.0
CCLIM3	1430	0	2.9	3.0	0.5	1.0	4.0	2.7 ~ 3.3
CCLIM4	1430	0	3.1	3.0	0.7	1.0	4.0	2.5 ~ 3.5
Factor 5*: Decision-making								
DEC1	1430	0	4.0	4.0	1.0	1.0	6.0	3.3 ~ 4.7
DEC2	1430	0	4.2	4.5	1.3	1.0	6.0	3.5 ~ 5.5
Factor 6*: Superior support								
SSUP1	1430	0	4.3	4.3	1.2	1.0	6.0	3.7 ~ 5.3
SSUP2	1430	0	4.4	4.5	1.3	1.0	6.0	3.5 ~ 5.5
Factor 7*: Peer support								
PSUP1	1430	0	4.6	4.7	1.0	1.0	6.0	4.0 ~ 5.3
PSUP2	1430	0	4.6	4.5	0.9	1.0	6.0	4.0 ~ 5.0
Factor 8‡: Self-esteem								
SELF1	1430	0	3.6	3.7	0.4	1.0	4.0	3.3 ~ 4.0
SELF2	1430	0	3.6	3.8	0.5	1.0	4.0	3.4 ~ 4.0
SELF3	1430	0	3.5	3.7	0.5	1.0	4.0	3.3 ~ 4.0
Factor 9‡: External locus of control								
ELC1	1430	0	2.9	3.0	0.6	1.0	4.8	2.6 ~ 3.4
ELC2	1430	0	3.0	3.0	0.6	1.0	5.0	2.5 ~ 3.5
ELC3	1430	0	2.8	2.8	0.5	1.0	4.8	2.4 ~ 3.2
ELC4	1430	0	2.2	2.2	0.6	1.0	4.5	1.8 ~ 2.5
ELC5	1430	0	2.5	2.4	0.6	1.0	4.8	2.0 ~ 3.0
Factor 10†: Emotional Exhaustion (high score means negative)								
EE1	1430	0	3.9	4.0	1.3	1.0	7.0	3.0 ~ 4.7
EE2	1430	0	3.5	3.3	1.3	1.0	7.0	2.7 ~ 4.3
EE3	1430	0	3.2	3.0	1.3	1.0	7.0	2.0 ~ 4.0
Factor 11†: Depersonalization (high score means negative)								
DP1	1430	0	2.3	2.0	1.1	1.0	6.7	1.3 ~ 3.0
DP2	1430	0	2.1	1.5	1.2	1.0	7.0	1.0 ~ 2.5
Factor 12†: Personal Accomplishment								

Table 1: Descriptive statistics for measurements *(continued)*

	n	n of NA	Central tendency		Dispersion tendency			
			Mean	Median	SD	Min	Max	Q1~Q3
PA1	1430	0	5.7	6.0	0.9	2.0	7.0	5.3 ~ 6.3
PA2	1430	0	5.8	6.0	1.0	2.0	7.0	5.5 ~ 6.5
PA3	1430	0	5.8	6.0	1.0	2.0	7.0	5.3 ~ 6.7

Note:

Indicators variables were formulated through item parcels.

[*] These indicators are parcels from Teacher Stress Scale instrument

[†] These indicators are parcels from BMI instrument

[‡] These parcels consist of items from single unidimensional scales

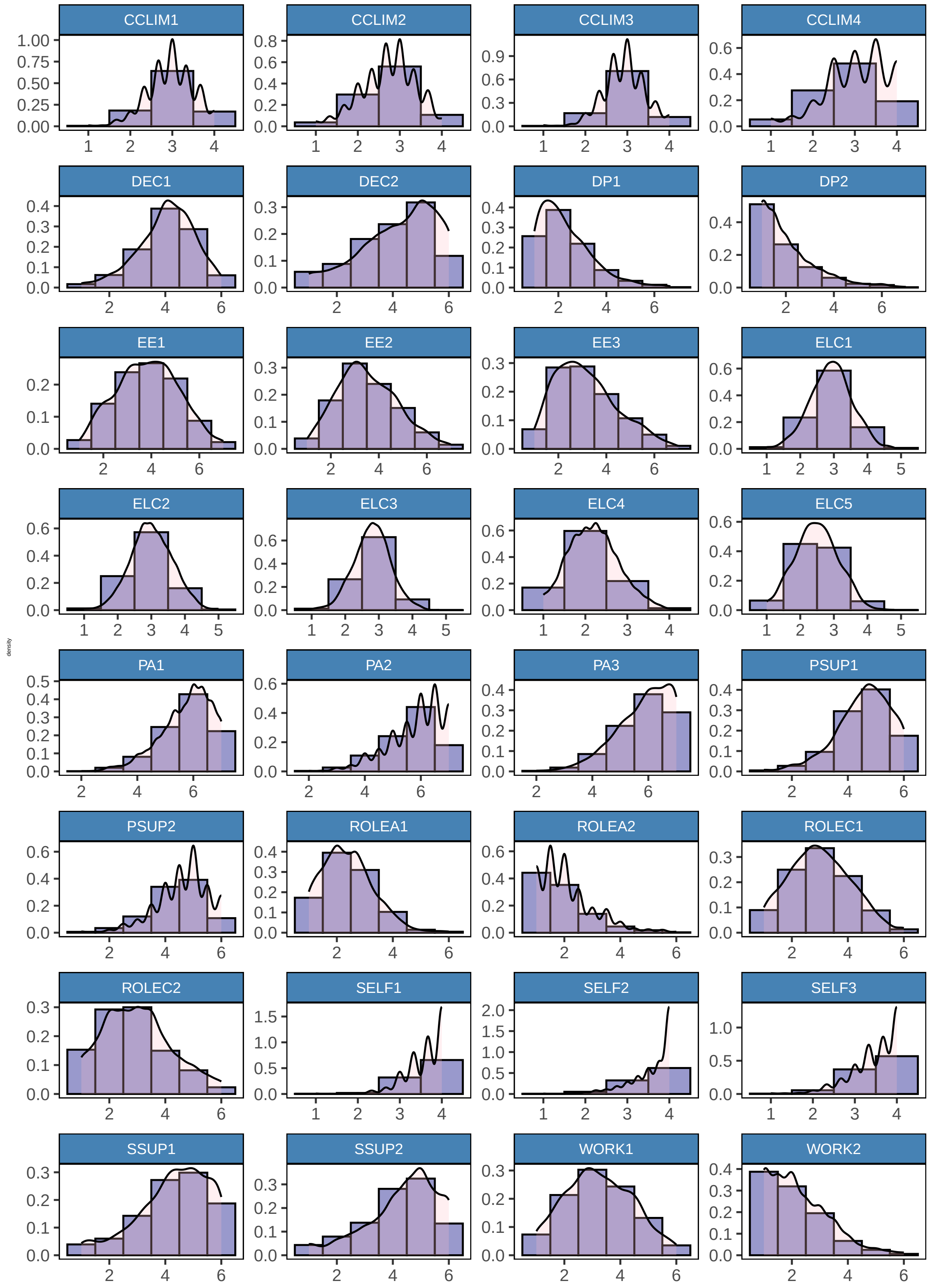
2.3.3 Visualization

(1) Histogram

Distribution of the data was examined via Histogram

```
corr.density(mbi, fig.num = 1)
```

Figure 1 Distribution of selected items



Ridge-line plots were generated for TSS and MBI indicators, respectively. By partially overlaying, it is a demonstration viable for comparing multiple distributions.

```
library(ggridges)
library(viridis)
library(hrbrthemes)
library(patchwork)
a <- mbi |>
  select(
    starts_with("EE")|starts_with("DP")|starts_with("PA")
    ) |>
  pivot_longer(everything(), names_to = "variable", values_to = "value") |>
  ggplot(aes(x = value, y = variable, fill = ..x..)) +
  geom_density_ridges_gradient(scale = 3, rel_min_height = 0.01) +
  scale_fill_viridis(name = "parcelled score", option = "C") +
  labs(title =
         'Fig2 (a). Distribution of indicator scores from BMI instrument') +
  labs(x = "Indicator scores", y = "Indicators") +
    theme(
      legend.position="none",
      panel.spacing = unit(0.1, "lines"),
      plot.title = element_text(size = 12),
      panel.grid.major = element_blank(),
      panel.background = element_rect(color = "black",
                                      fill = "white")
      )

b <- mbi |>
  select(
    starts_with("ROL")|starts_with("WOR")|starts_with("DEC")|contains("SUP")
    ) |>
  pivot_longer(everything(), names_to = "variable", values_to = "value") |>
  ggplot(aes(x = value, y = variable, fill = ..x..)) +
  geom_density_ridges_gradient(scale = 3, rel_min_height = 0.01) +
  scale_fill_viridis(name = "parcelled score", option = "C") +
  labs(title =
         'Fig2 (b). Distribution of indicator scores from TSS instrument') +
  labs(x = "Indicator scores", y = "Indicators")+
    theme(
      legend.position="none",
      panel.spacing = unit(0.1, "lines"),
      plot.title = element_text(size = 12),
      panel.grid.major = element_blank(),
      panel.background = element_rect(color = "black",
                                      fill = "white")
    )

a/b
```

Fig2 (a). Distribution of indicator scores from BMI instrument

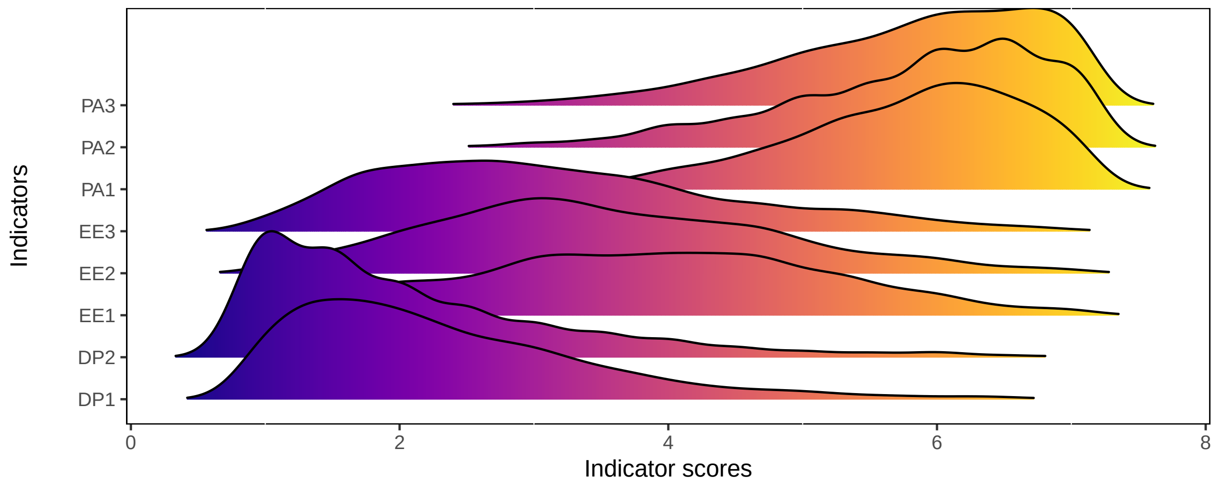
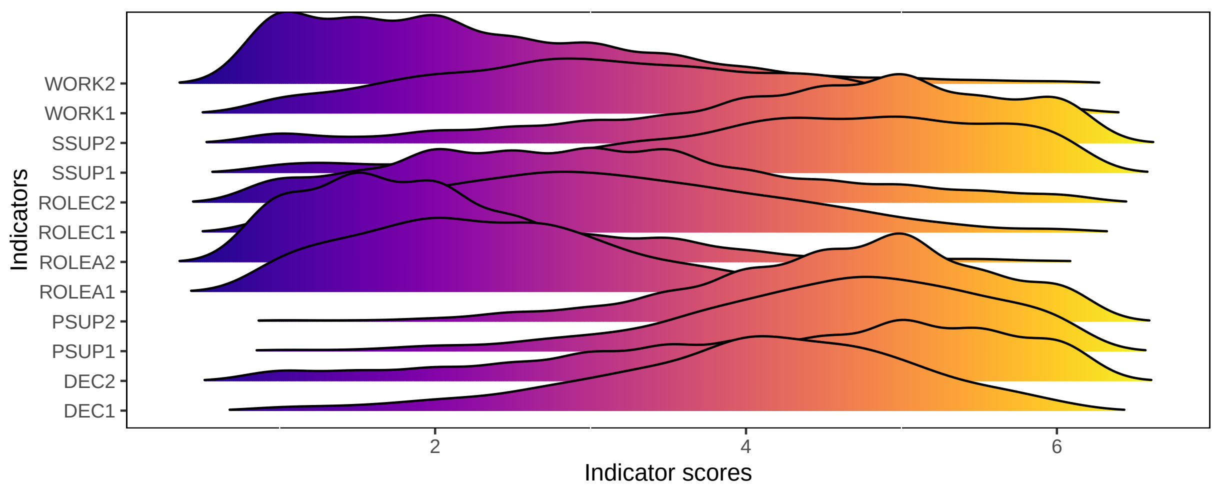


Fig2 (b). Distribution of indicator scores from TSS instrument

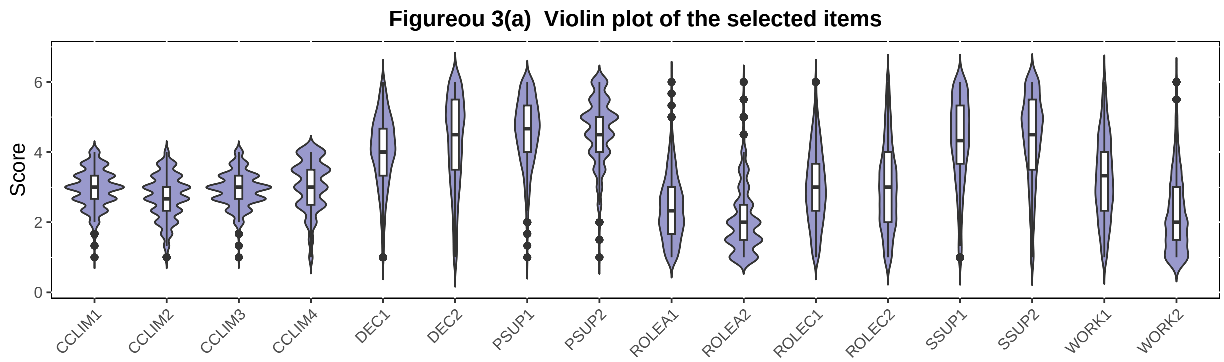


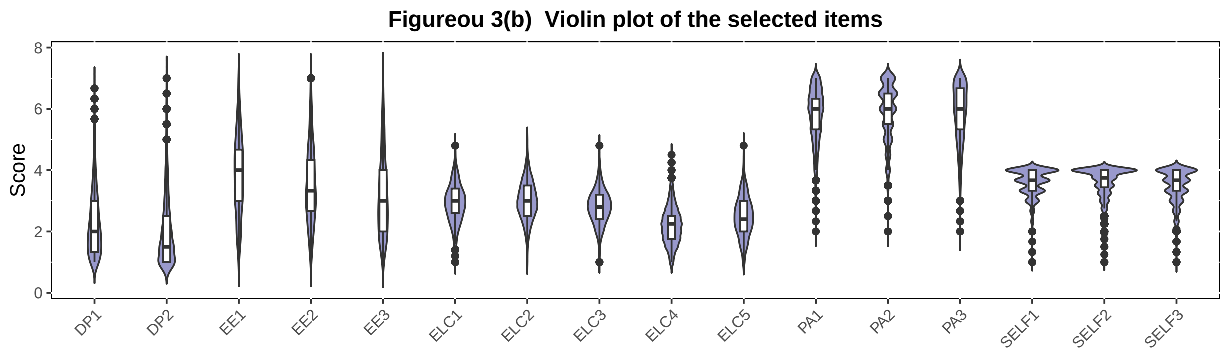
Clearly, within each instrument, indicators for same factor tend to show similar distribution features.

(2) Violin plot

Violin plot also provides information on distribution, plus ideas on out-liers.

```
a <- mbi |>
  select(1:16) |>
  violin.box(fig.num = "3(a)")

b <- mbi |>
  select(17:32) |>
  violin.box(fig.num = "3(b)")
library(patchwork)
a/b
```

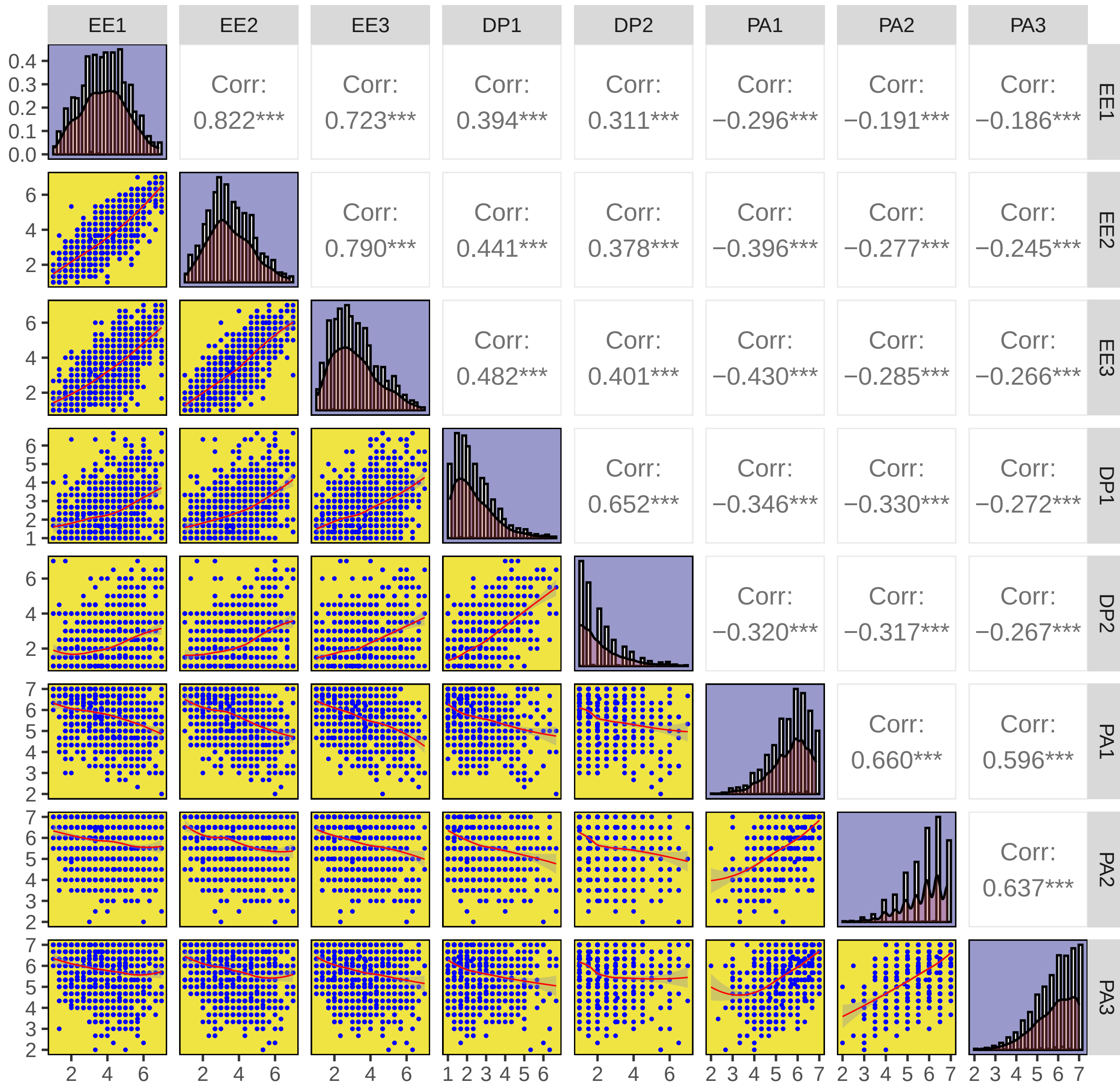




(3) Correlation among items

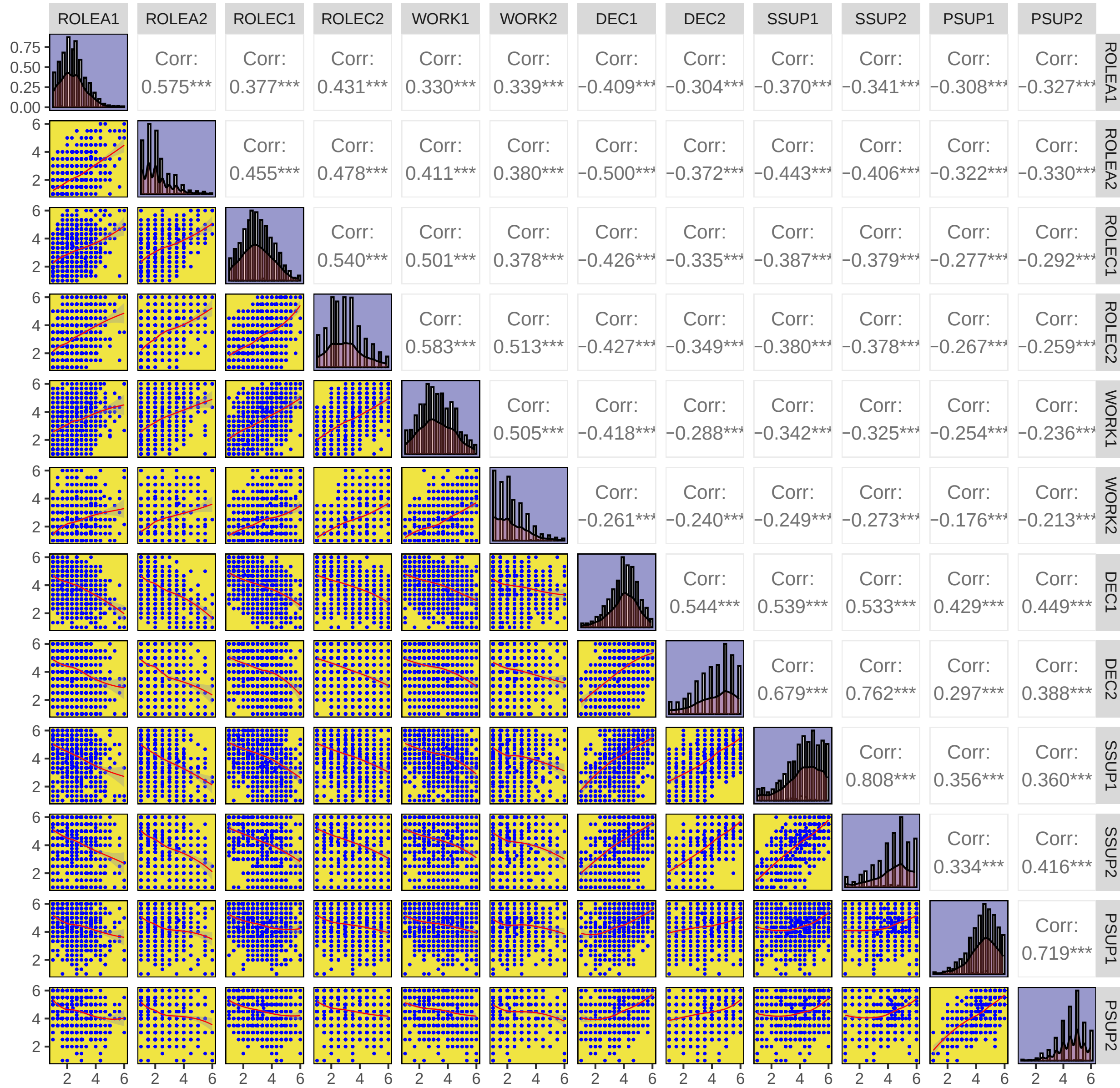
```
#draw it
mbi |>  select(starts_with("EE")|starts_with("DP")|starts_with("PA")) |>
  ggpairs(lower =
            list(continuous = my.fun.smooth), #lower half show points with fitted line
          diag =
            list(continuous = my.fun.density), #diagonal grids show density plots
          title = "Fig4(a). Relationships among parcels of MBI instrument") + #title
  theme (plot.title = element_text(size = 12,  #adjust title visuals
                                   face = "bold"))
```

Fig4(a). Relationships among parcels of MBI instrument




```
mbi |>  select(starts_with("ROL")|starts_with("WOR")|starts_with("DEC")|contains("SUP")) |>
  ggpairs(lower =
          list(continuous = my.fun.smooth), #lower half show points with fitted line
        diag =
          list(continuous = my.fun.density), #diagonal grids show density plots
        title = "Fig4(b). Relationships among parcels of TSS instrument") + #title
  theme (plot.title = element_text(size = 14,  #adjust title visuals
                                   face = "bold"))
```

Fig4(b). Relationships among parcels of TSS instrument



3 Testing the for the validity of causal structure of burnout

3.1 Initial full structural equation model (hypothesized model modified according to CFA)

This full structural equation model was a hypothesized model. I have established causal structure linking several stressor variables considered to contribute to the presence of burnout (Fig. 4). These postulated causal relations linking variables were supported in theory and/or empirical research. I wanted to test the hypothesis that the causal pattern was true. The findings would contribute to the understanding of key determinants of teacher burnout. Since the hypothesis was proven not true, I performed post-hoc analysis to improve the causal structure step by step, until best fitting, albeit most parsimonious, model of any set of tested models had been achieved.

Note that *"an important preliminary step in the analysis of full latent variable models is to test first for the validity of the measurement model before making any attempt to evaluate the structural model. Accordingly, CFA procedures are used in testing the validity of the indicator variables. Once it is known that the measurement model is operating adequately, one can then have more confidence in findings related to assessment of the hypothesized structural model."* The current analysis started at the point where CFA had already been done. As described by Byrne, the analysis produced fit indices showing exceptionally good fit to the data; nonetheless, CFA model for the TSS was re-specified to include two additional parameters, both about allowing for cross-loading terms (DEC2 cross-loads onto Factor 1; DEC2 cross-loads onto Factor 5). These parameters set free were incorporated into the initial hypothesized model (Fig. 4). ### Define the initial model

```
library(semPlot)#install.packages("semPlot")

initial_model <- '
# Factors:
 F1ROLA =~ ROLEA1 + ROLEA2 + DEC2
 F2ROLC =~ ROLEC1 + ROLEC2
 F3WORK =~ WORK1 + WORK2
 F4CLIM =~ CCLIM1 + CCLIM2 + CCLIM3 + CCLIM4
 F5DEC =~ DEC1 + DEC2
 F6SSUP =~ SSUP1 + SSUP2 + DEC2
 F7PSUP =~ PSUP1 + PSUP2
 F8SELF =~ SELF1 + SELF2 + SELF3
 F9ELC =~ ELC1 + ELC2 + ELC3 + ELC4 + ELC5
 F10EE =~ EE1 + EE2 + EE3
 F11DP =~ DP1 + DP2
 F12PA =~ PA1 + PA2 + PA3
# Regressions:
 F8SELF ~ F5DEC + F6SSUP + F7PSUP
 F9ELC ~ F5DEC
 F10EE ~ F2ROLC + F3WORK + F4CLIM
 F11DP ~ F2ROLC + F10EE
 F12PA ~ F1ROLA + F8SELF + F9ELC + F10EE + F11DP
'
```

3.1.1 Visualize the initial model

To approximate the visual effect on slides, the coordinates for each nodes were defined on a 60 by 72 matrix.

```
#generate a matrix
m <- matrix(NA, 60, 72)

#define positions of the factors
m[12, 68] <- "F1ROLA"
m[12, 40] <- "F2ROLC"
m[12, 28] <- "F3WORK"
m[12,12] <- "F4CLIM"
m[21,12] <-"F5DEC"
m[40,12] <-"F6SSUP"
m[53,9] <-"F7PSUP"
m[44,24] <-"F8SELF"
m[52,40] <-"F9ELC"
m[37,48] <-"F10EE"
```

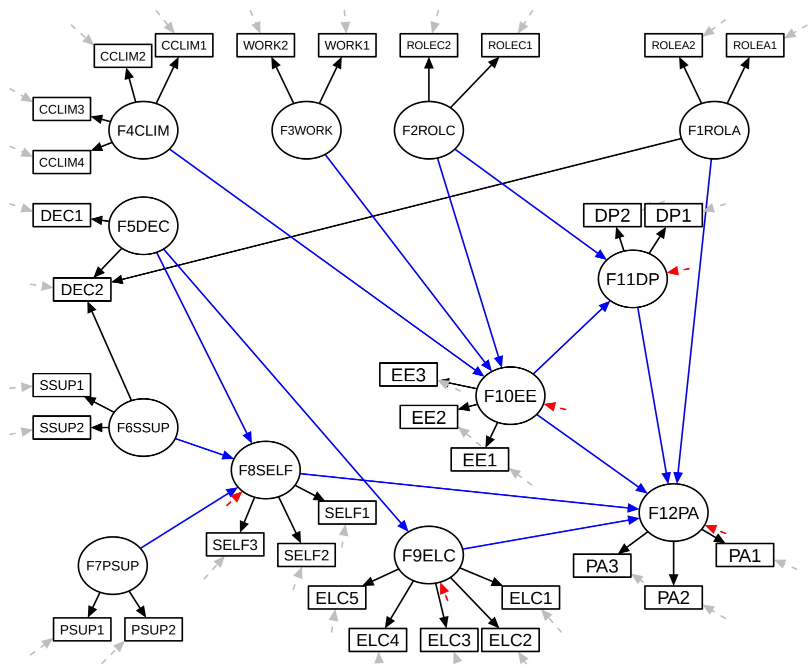
```r
m[26,60] <-"F11DP"
m[48,64] <-"F12PA"

#define the posiitions of the indicators (parcelled items)
m[4, 72] <- "ROLEA1"
m[4, 64] <- "ROLEA2"
m[4, 48] <- "ROLEC1"
m[4, 40] <- "ROLEC2"
m[4, 32] <- "WORK1"
m[4, 24] <- "WORK2"
m[4, 16] <- "CCLIM1"
m[5, 10] <- "CCLIM2"
m[10, 4] <- "CCLIM3"
m[15, 4] <- "CCLIM4"
m[20, 4] <- "DEC1"
m[27, 6] <- "DEC2"
m[36, 4] <- "SSUP1"
m[40, 4] <- "SSUP2"
m[59, 6] <- "PSUP1"
m[59, 13] <- "PSUP2"
m[48, 32] <- "SELF1"
m[52, 28] <- "SELF2"
m[51, 21] <- "SELF3"
m[56, 50] <- "ELC1"
m[60, 48] <- "ELC2"
m[60, 42] <- "ELC3"
m[60, 35] <- "ELC4"
m[56, 31] <- "ELC5"
m[43, 45] <- "EE1"
m[39, 40] <- "EE2"
m[35, 38] <- "EE3"
m[20, 64] <- "DP1"
m[20, 58] <- "DP2"
m[52, 71] <- "PA1"
m[56, 64] <- "PA2"
m[53, 57] <- "PA3"
```

The diagram of the initial model was generated.

```r
semPaths(semPlotModel(initial_model),
             style = "lisrel",
             rotation = 2,
             sizeLat = 6,
             sizeLat2 = 5,
             sizeMan = 5,
             sizeMan2 = 2,
             residScale = 4,
             shapeMan = "rectangle",
             edge.color = c(rep("black",34),
                            rep("blue",14),
                            rep("gray",32),
                            rep("red",5)),
         residuals = T,
```

```
        layout = m,
        nCharNodes=0,
        optimizeLatRes = T,
        exoVar = F)
title(main = list("Figure 5. Hypothesized model of teacher burnout",
                  cex = 1.5, font =1),
      outer = F, line = -1)
title(sub = "Notes: Red arrow indicates factor residuals; gray arrow indicates error residuals;
     blue arrow indicates regression path; black arrow indicates factor loading",
     line = 0, adj = 0.7)
```

Figure 5. Hypothesized model of teacher burnout



Notes: Red arrow indicates factor residuals; gray arrow indicates error residuals;
blue arrow indicates regression path; black arrow indicates factor loading

3.1.2 Estimate the SEM model (initial)

```
library(lavaan)
model1 <- initial_model # defined above

# Estimate the model with the robust (MLM) estimator:
sem1 <- sem(model1, data = mbi, estimator = "MLM", mimic = "Mplus")

# Numerical summary of the model:
sem.summary.mlm.a(sem1, 12, 32, "mlm", "Model fit indices for initial model")
```

Table 3: Residual variance of structural regression path and select factors for model1

Parameter*	B†	Beta‡	SE	Z	p-value
Strutural regression path					
F5DEC→F8SELF	0.475	0.990	0.050	9.462	0
F6SSUP→F8SELF	-0.155	-0.490	0.025	-6.194	0
F7PSUP→F8SELF	-0.066	-0.150	0.029	-2.307	0.021
F5DEC→F9ELC	-0.288	-0.476	0.021	-13.784	0
F2ROLC→F10EE	-8.707	-5.765	6.310	-1.380	0.168
F3WORK→F10EE	8.082	6.325	5.321	1.519	0.129
F4CLIM→F10EE	-0.93	-0.272	0.658	-1.414	0.157
F2ROLC→F11DP	0.258	0.203	0.048	5.356	0
F10EE→F11DP	0.373	0.444	0.033	11.377	0
F1ROLA→F12PA	-0.071	-0.062	0.046	-1.548	0.122
F8SELF→F12PA	0.472	0.217	0.085	5.562	0
F9ELC→F12PA	-0.208	-0.121	0.049	-4.264	0
F10EE→F12PA	-0.064	-0.098	0.025	-2.564	0.01
F11DP→F12PA	-0.218	-0.281	0.031	-7.115	0
Selected factors					
F8SELF	0.079	0.682	0.007	10.672	0
F9ELC	0.143	0.774	0.011	13.432	0
F10EE	-0.432	-0.332	0.770	-0.562	0.574
F11DP	0.605	0.658	0.047	12.880	0
F12PA	0.383	0.695	0.023	16.815	0

Note:
Values highlighted in red should be taken note of
* → indicates regression path
† Crude estimates
‡ Standardized estimates

Table 2: Model fit indices for initial model

Measure	Value
chi square	1554.942
df	427.000
p value	0.000
CFI	0.945
TLI	0.936
RMSEA	0.043
RMSEA p value	1.000
SRMR	0.051
CSF	1.117

```
#summary(sem1, fit.measures = TRUE, standardized = TRUE)
```

```
#print concern table for model 1
concern.table(sem1, 14, "model1")
```

3.1.3 Comments on the result (initial model)

Here we see that the re-scaled 2 value (i.e., the MLM 2) is 1541.844 with 427 degrees of freedom. The reported chi square scaling factor value for the MLM estimator indicates that if the MLM 2 were multiplied by 1.127, it would approximate the uncorrected ML 2 value (1737.658).

Given the large number of parameters estimated in this model, the reported results are understandably lengthy. Thus, in the interest of space, I report findings pertinent to only the structural parameters, as well as a few residual variances. These results are presented in Table 3..

Examination of estimated parameters in the model revealed all to be statistically significant except for those highlighted in red in Figure 3. These non-significant parameters the following structural regression paths: (a) F10 on F2 (Role Conflict → Emotional Exhaustion), F10 on F3 (Work Overload → Emotional Exhaustion), F10 on F4 (Classroom Climate → Emotional Exhaustion), and (b) F12 on F1 (Role Ambiguity → Personal Accomplishment).

For the negative variance for the residual associated with Factor 10 (highlighted in red in table 3), I left it as it was for the time being.

3.1.4 Model mis-specification

A review of the MIs reveals some evidence of misfit in the model. Because we are interested solely in the causal paths of the model at this point, only MIs related to these parameters are included in table 5. The sense-making for the choice of parameter to set free was also placed in the table.

```
#extract needed variables
MI.model1 <- modindices(sem1,
                 standardized = TRUE,
                 sort. = TRUE,
                 maximum.number = 20) |>
  filter(op == "~") |>
  filter(lhs %in%  #When these factors are predicted variables, it is related
           c("F8SELF",  # to the topic of this study
             "F9ELC",
             "F10EE",
             "F11DP",
             "F12PA"))

#adapt to publication style
MI.model1 <- MI.model1 |>
  mutate(Parameter = paste(rhs, "→", lhs)) |>
  select(Parameter, MI = mi, EPC = epc, "std EPC" = sepc.all) |>
  filter(MI>10) |> #for saving space, the number of parameters was managed
  mutate("Logics" = c("Sensible and meaningful",
                      "illogical (wrong direction of correlation)",
                      "Sensibe but not meaningful for this study"))

#add footnote symbol to parameters
  for (i in 1:nrow(MI.model1)){
  symbol <- c("*", "†", "‡") #symbols for footnotes
  Parameter <- unlist(MI.model1$Parameter)
  MI.model1$Parameter[i] <-  paste0(Parameter[i], symbol[i])
  }

#Visualize the table
```

Table 4: Selected modification indices for initial model

Parameter	MI	EPC	std EPC	Logics
F4CLIM → F11DP*	112.597	-0.974	-0.339	Sensible and meaningful
F3WORK → F11DP†	85.318	-4.799	-4.467	illogical (wrong direction of correlation)
F12PA → F8SELF‡	49.986	0.329	0.719	Sensibe but not meaningful for this study

Note:
Parameter highlighted in red is selected for modification
* Poorer Classroom climate leads to worsening depersonalization
† Less workload leads to worsening depersonalization
‡ Higher accomplishment results in increased self-esteem

```
MI.model1 |>
  kable(digits = 3,
        booktab = T,
        linesep = "",
        caption = "Selected modification indices for initial model") |>
  kable_styling(latex_options = "striped") |>
  row_spec(1, color = "red") |> #this is the row with parameter to set free
  footnote(general =
             "Parameter highlighted in red is selected for modification",
    symbol = c("Poorer Classroom climate leads to worsening depersonalization",
             "Less workload leads to worsening depersonalization",
             "Higher accomplishment results in increased self-esteem"))
```

3.2 Post hos analysis (Model 2)

3.2.1 Compare SEM model 2 with initial model

I defined model 2 as per the conclusion from last section: set the regression path leading from F4 to F11 (Poorer Classroom climate leads to worsening depersonalization) free to estimate (Fig 6). The model fit indices and its comparison with the preceding model (model1) was tabulated in table 5.

```
model2 <- paste(initial_model, "F11DP ~ F4CLIM")

# Estimate the model with the robust (MLM) estimator:
sem2 <- sem(model2, data = mbi, estimator = "MLM", mimic = "Mplus")
```

```
#extract needed fit indices in model1
sem.measure1 <- fitMeasures(sem1,    #obtain specified measured.
                            c("chisq.scaled",
                              "df.scaled",
                              "pvalue.scaled",
                              "cfi.scaled",
                              "tli.scaled",
                              "rmsea.scaled",
                              "srmr_bentler",
                              "chisq.scaling.factor")) |>
  t() |>
```

Table 5: Comparison of new and preceding models

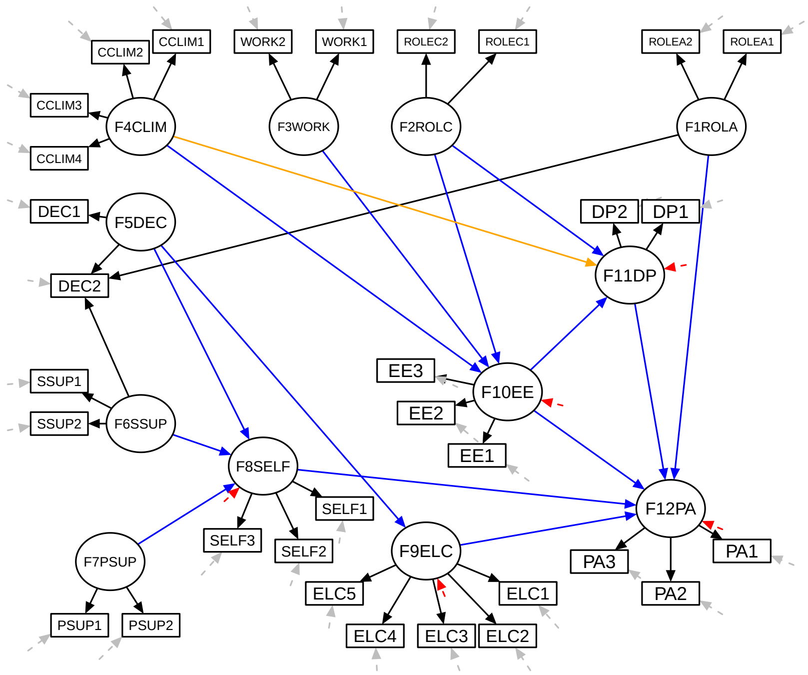
	Chi square	DF	p value	ΔChi square	CFI	TLI	RMSEA	SRMR	CSF
model1	1554.942	427	0	–	0.945	0.936	0.043	0.051	1.117
model2*	1450.985	426	0	91.67	0.950	0.941	0.041	0.046	1.117

* Model1 + parameter 'F4→F11' set free to estimate

```
  round(3)
#extract needed fit indices in model2
sem.measure2 <- fitMeasures(sem2,    #obtain specified measured.
                            c("chisq.scaled",
                              "df.scaled",
                              "pvalue.scaled",
                              "cfi.scaled",
                              "tli.scaled",
                              "rmsea.scaled",
                              "srmr_bentler",
                              "chisq.scaling.factor")) |>
  t() |>
  round(3)
#combine the 2 sets of indices
sem.compare2 <- rbind(sem.measure1, sem.measure2) |> data.frame()

#add column names
colnames(sem.compare2) <- c("Chi square", "DF", "p value", "CFI", "TLI",
                            "RMSEA", "SRMR", "CSF")
#turn named vector to data frame and pass into a new object
sem.compare2.tab<- sem.compare2 %>%
  mutate("ΔChi square" = chi.diff(sem1, sem2)) |>
  select("Chi square", "DF", "p value", "ΔChi square", "CFI", "TLI", "RMSEA", "SRMR", "CSF")

#the first model does not have chisq diff value, so place a "--" in the cell
sem.compare2.tab[1,4] <- "--"
rownames(sem.compare2.tab) <- c("model1", "model2*")

#aesthetics fine-tune and print the table
sem.compare2.tab |>
  kable(booktab =T,
        #format = "markdown",
        caption = "Comparison of new and preceding models",
        align = "r") |>
  kable_styling() |>
  footnote(symbol = "Model1 + parameter 'F4→F11' set free to estimate")
```

```
#plot the diagram
full.sem.diagram(model2, 15, "6", "model2")
```

Figure 6 Modified model (model2 of teacher burnout)



Notes: Orange arrow is the regression path set free to estimate in the current model

3.2.2 Examine the parameters of concern for model 2

```
#regression path estimates
concern.table(sem2, 15, "model2")
```

3.2.3 Comment on the result (model 2)

The estimation of Model 2 yielded an overall MLM 2(426) value of 1450.985 (scaling correction factor = 1.117); CFI and TLI values of 0.950 and 0.945, respectively; a RMSEA value of 0.041; and a SRMR value of 0.046 (Table 5). Although improvement in model fit for Model 2, compared with the originally hypothesized model, would appear to be somewhat minimal on the basis of the CFI, TLI, RMSEA, and SRMR values, the corrected chi-square difference test was found to be significant (MLM Δ 2[1] = 91.67), which finalized the decision (Table 5).

Notes from Byrne's book: "the thrust of these post-hoc analyses is to fine-tune our hypothesized structure such that it includes all viable and statistically significant structural paths, and, at the same time, eliminates all non-significant paths. Consequently, as long as the 2-difference test is statistically significant, and the newly added parameters are substantively meaningful, I consider the post-hoc analyses to be appropriate."

The anomaly of negative residual variance remained in the output for Model 2 (Table 6). The estimated structural regression paths for the three factors hypothesized to influence Factor 10 (Factors 2, 3, and 4) and F1 to influence Factor 12 remained statistically non-significant (Table 8).

Besides, The negative residual variance of F10 remained as it was in model 2.

Table 6: Residual variance of structural regression path and select factors for model2

Parameter*	B†	Beta‡	SE	Z	p-value
Strutural regression path					
F5DEC→F8SELF	0.474	0.989	0.050	9.453	0
F6SSUP→F8SELF	-0.155	-0.489	0.025	-6.198	0
F7PSUP→F8SELF	-0.066	-0.150	0.029	-2.308	0.021
F5DEC→F9ELC	-0.288	-0.476	0.021	-13.791	0
F2ROLC→F10EE	-8.396	-5.553	5.945	-1.412	0.158
F3WORK→F10EE	7.856	6.151	5.034	1.561	0.119
F4CLIM→F10EE	-0.563	-0.164	0.633	-0.888	0.374
F2ROLC→F11DP	0.173	0.135	0.046	3.726	0
F10EE→F11DP	0.299	0.354	0.031	9.627	0
F1ROLA→F12PA	-0.069	-0.060	0.046	-1.489	0.137
F8SELF→F12PA	0.473	0.217	0.084	5.599	0
F9ELC→F12PA	-0.204	-0.118	0.049	-4.190	0
F10EE→F12PA	-0.058	-0.089	0.024	-2.411	0.016
F11DP→F12PA	-0.228	-0.297	0.030	-7.533	0
F4CLIM→F11DP	-0.969	-0.334	0.100	-9.642	0
Selected factors					
F8SELF	0.079	0.682	0.007	10.672	0
F9ELC	0.143	0.774	0.011	13.432	0
F10EE	-0.432	-0.332	0.770	-0.562	0.574
F11DP	0.605	0.658	0.047	12.880	0
F12PA	0.383	0.695	0.023	16.815	0

Note:

Values highlighted in red should be taken note of

* → indicates regression path

† Crude estimates

‡ Standardized estimates

3.2.4 Model mis-specification (model 2)

```
#extract needed variables
MI.model2 <- modindices(sem2,
                  standardized = TRUE,
                  sort. = TRUE,
                  maximum.number = 25) |>
  filter(op == "~") |>
  filter(lhs %in%  #When these factors are predicted variables, it is related
           c("F8SELF",  # to the topic of this study
             "F9ELC",
             "F10EE",
             "F11DP",
             "F12PA"))

#adapt to publication style
MI.model2 <- MI.model2 |>
  mutate(Parameter = paste(rhs, "→", lhs)) |>
  select(Parameter, MI = mi, EPC = epc, "std EPC" = sepc.all) |>
  filter(MI>10) |>
  mutate("Logics" = c("Sensible but not meaningful for this study",
                      "Sensible and meaningful",
                      "Sensibe and meaningful, but MI is lower than F5→F12"))


#add footnote symbol to parameters
  for (i in 1:nrow(MI.model2)){
  symbol <- c("*", "†", "‡")
  Parameter <- unlist(MI.model2$Parameter)
  MI.model2$Parameter[i] <-  paste0(Parameter[i], symbol[i])
  }

#Visualize the table
MI.model2 |>
  kable(digits = 3,
        booktab = T,
        linesep = "",
        caption = "Selected modification indices for initial model") |>
  kable_styling(latex_options = "striped") |>
  row_spec(2, color = "red") |>
  footnote(general =
             "Parameter highlighted in red is selected for modification",
    symbol = c("Higher accomplishment leads to increased self-esteem",
             "Invovlement of more decision making gives sense of accomplishment",
             "People with high esteem will less likely get emotionally exhausted"))
```

3.3 Post hos analysis (Model 3)

3.3.1 Compare SEM model 3 with preceding models

I defined model 3 as per the conclusion from last section: on the basis of model 2, I set the regression path leading from F5 to F12 (Invovlement of more decision making gives sense of accomplishment) free to estimate

Table 7: Selected modification indices for initial model

Parameter	MI	EPC	std EPC	Logics
F12PA → F8SELF*	49.859	0.330	0.720	Sensible but not meaningful for this study
F5DEC → F12PA†	47.088	0.480	0.458	Sensible and meaningful
F8SELF → F10EE‡	42.770	-1.029	-0.306	Sensibe and meaningful, but MI is lower than F5→F12

Note:

Parameter highlighted in red is selected for modification

* Higher accomplishment leads to increased self-esteem

† Invovlement of more decision making gives sense of accomplishment

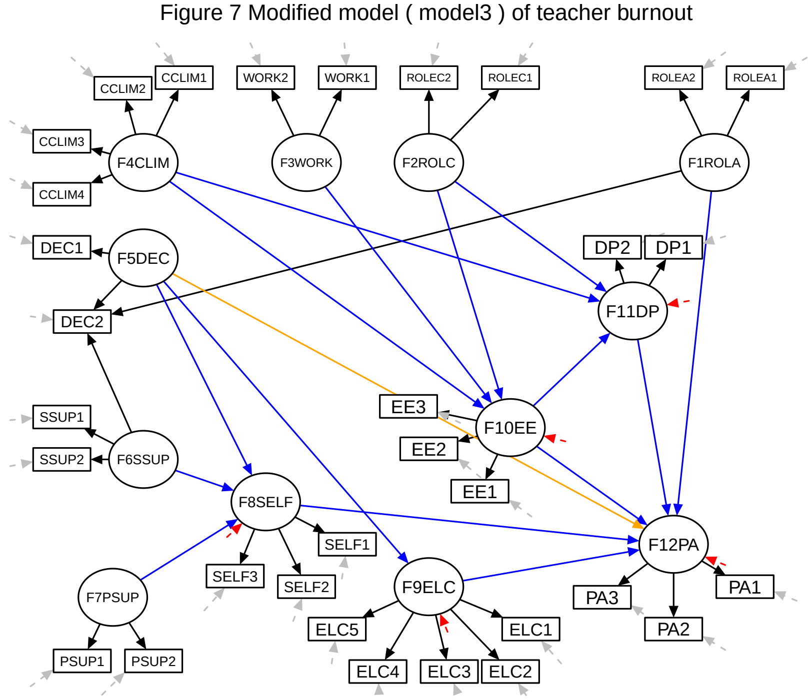
‡ People with high esteem will less likely get emotionally exhausted

(Fig 7). The model fit indices and their comparison with the preceding models were tabulated in table 8.

```
#add new freely estimated parameter to the preceding model
model3 <- paste(model2, "\nF12PA ~ F5DEC")

# Estimate the model with the robust (MLM) estimator:
sem3 <- sem(model3, data = mbi, estimator = "MLM", mimic = "Mplus")
```

```
full.sem.diagram(model3, 16, "7", "model3")
```

Figure 7 Modified model (model3) of teacher burnout



Notes: Orange arrow is the regression path set free to estimate in the current model

```r
sem.measure3 <- fitMeasures(sem3,    #obtain specified measured.
                            c("chisq.scaled",
                              "df.scaled",
                              "pvalue.scaled",
                              "cfi.scaled",
                              "tli.scaled",
                              "rmsea.scaled",
                              "srmr_bentler",
                              "chisq.scaling.factor")) |>
  t() |>
  round(3)
#combine the 2 sets of indices and save into a new object
sem.compare3 <- rbind(sem.measure1, sem.measure2, sem.measure3) |> data.frame()
sem.compare3.tab <- sem.compare3

#add column "chi square" and place values into it
sem.compare3.tab$"ΔChi-square" <- rep(NA, 3)
sem.compare3.tab$"ΔChi-square"[2] <- chi.diff(sem1, sem2)#calculate chisq difference
sem.compare3.tab$"ΔChi-square"[3] <- chi.diff(sem2, sem3)

#turn named vector to data frame
sem.compare3.tab<- sem.compare3.tab %>%
  select("Chi-square" = chisq.scaled, "DF"=df.scaled,
         "p value"=pvalue.scaled, "ΔChi-square", "CFI"= cfi.scaled,
         "TLI"= tli.scaled, "RMSEA"= rmsea.scaled,
         "SRMR"=srmr_bentler, "CSF"=chisq.scaling.factor)

#There is no chisq-diff value for the first model, so place a "--" in the cell
sem.compare3.tab[1,4] <- "--"

#add row names
rownames(sem.compare3.tab) <- c("model1", "model2*", "Model3†")

#table aesthetics finetune and print table
sem.compare3.tab |>
  kable(booktab =T,
        #format = "markdown",
        caption = "Comparison of new and preceding models",
        align = "r") |>
  kable_styling() |>
  footnote(symbol = c("Model1 +  parameter 'F4→F11' set free to estimate",
                      "Model2 +  parameter 'F5→F12' set free to estimate")
           )
```

3.3.2 Examine the parameters of concern for model 3

```r
concern.table(sem3, 16, "model3")
```

Table 8: Comparison of new and preceding models

	Chi-square	DF	p value	ΔChi-square	CFI	TLI	RMSEA	SRMR	CSF
model1	1554.942	427	0	–	0.945	0.936	0.043	0.051	1.117
model2*	1450.985	426	0	91.67	0.950	0.941	0.041	0.046	1.117
Model3†	1406.517	425	0	46.866	0.952	0.944	0.040	0.044	1.117

* Model1 + parameter 'F4→F11' set free to estimate

† Model2 + parameter 'F5→F12' set free to estimate

Table 9: Residual variance of structural regression path and select factors for model3

Parameter*	B†	Beta‡	SE	Z	p-value
Strutural regression path					
F5DEC→F8SELF	0.473	0.989	0.050	9.432	0
F6SSUP→F8SELF	-0.157	-0.496	0.025	-6.363	0
F7PSUP→F8SELF	-0.061	-0.138	0.028	-2.197	0.028
F5DEC→F9ELC	-0.288	-0.477	0.021	-13.722	0
F2ROLC→F10EE	-6.345	-4.198	3.393	-1.870	0.062
F3WORK→F10EE	6.113	4.797	2.870	2.130	0.033
F4CLIM→F10EE	-0.576	-0.168	0.488	-1.181	0.237
F2ROLC→F11DP	0.176	0.137	0.046	3.779	0
F10EE→F11DP	0.3	0.353	0.031	9.611	0
F1ROLA→F12PA	0.333	0.285	0.085	3.917	0
F8SELF→F12PA	0.469	0.213	0.085	5.488	0
F9ELC→F12PA	-0.153	-0.088	0.053	-2.919	0.004
F10EE→F12PA	0.025	0.038	0.028	0.887	0.375
F11DP→F12PA	-0.2	-0.259	0.030	-6.575	0
F4CLIM→F11DP	-0.963	-0.330	0.100	-9.599	0
F5DEC→F12PA	0.504	0.480	0.083	6.054	0
Selected factors					
F8SELF	0.079	0.682	0.007	10.672	0
F9ELC	0.143	0.774	0.011	13.432	0
F10EE	-0.432	-0.332	0.770	-0.562	0.574
F11DP	0.605	0.658	0.047	12.880	0
F12PA	0.383	0.695	0.023	16.815	0

Note:

Values highlighted in red should be taken note of

* → indicates regression path

† Crude estimates

‡ Standardized estimates

3.3.3 Comment on the result (model 3)

Model 3 yielded an overall MLM 2(425) value of 1406.517 (scaling correction factor = 1.117), with CFI = 0.952, TLI = 0.944, RMSEA = 0.040, and SRMR = 0.044. Again, the MLM 2 difference between Models 2 and 3 is statistically significant (MLM 2[1] = 46.866), albeit differences in the other fit indices across Models 2 and 3 were once again minimal. See Tab 8.

As expected, the estimate for the newly incorporated path from Decision Making to Personal Accomplishment (F5 → F12) was found to be statistically significant. However, the three previous non-significant path remained p values>0.05 (two leading to F12; one leading to F10). Besides, The negative residual variance of F10 remained as it was in model 3 (see tab 9, estimate highlighted in red).

3.3.4 Model mis-specification (model 3)

```r
#extract needed variables
MI.model3 <- modindices(sem3,
                  standardized = TRUE,
                  sort. = TRUE,
                  maximum.number = 50) |>
  filter(op == "~") |>
  filter(lhs %in%  #When these factors are predicted variables, it is related
           c("F8SELF",  # to the topic of this study
             "F9ELC",
             "F10EE",
             "F11DP",
             "F12PA"))

#adapt to publication style
MI.model3 <- MI.model3 |>
  mutate(Parameter = paste(rhs, "→", lhs)) |>
  select(Parameter, MI = mi, EPC = epc, "std EPC" = sepc.all) |>
  filter(MI>10) |>
  head(3) |>
  mutate("Logics" = c("Sensible and meaningful, but not as sensible as F2→F9",
                      "Highly sensible and meaningful",
                      "Sensible and meaningful, but MI < other 2"))


#add footnote symbol to parameters
  for (i in 1:nrow(MI.model3)){
  symbol <- c("*", "†", "‡")
  Parameter <- unlist(MI.model3$Parameter)
  MI.model3$Parameter[i] <-  paste0(Parameter[i], symbol[i])
  }

#Visualize the table
MI.model3 |>
  kable(digits = 3,
        booktab = T,
        linesep = "",
        caption = "Selected modification indices for initial model") |>
  kable_styling(latex_options = "striped") |>
  row_spec(2, color = "red") |>
```

Table 10: Selected modification indices for initial model

Parameter	MI	EPC	std EPC	Logics
F8SELF → F10EE*	43.412	-1.021	-0.304	Sensible and meaningful, but not as sensible as F2→F9
F2ROLC → F9ELC†	41.343	0.191	0.336	Highly sensible and meaningful
F8SELF → F9ELC‡	39.841	-0.284	-0.225	Sensible and meaningful, but MI < other 2

Note:

Parameter highlighted in red is selected for modification

[*] Lower self-esteem will causes higher emotional exhaustion

[†] Higher role conflict increases external locus of control

[‡] Lower self-esteem leads to higher external locus of control

```
footnote(general =
           "Parameter highlighted in red is selected for modification",
   symbol = c("Lower self-esteem will causes higher emotional exhaustion",
           "Higher role conflict increases external locus of control",
           "Lower self-esteem leads to higher external locus of control"))
```

Quoted from Byrne, "Again, I believe it is worthwhile to note why two alternate MI values, close in value to the one chosen here, are considered to be inappropriate. I refer to results related to the structural paths of F10 on F8 (MI = 38.868) and of F9 on F8 (MI = 35.670). In both cases, the flow of causal direction is incorrect." I don't agree with her comments on wrong causal direction. To me, it is fairly logical, though not perfectly straightforward, that Lower self-esteem will cause higher emotional exhaustion (F8→F10, negative sign of EPC), and that Lower self-esteem leads to higher external locus of control (F8→F9, negative sign of EPC). See table 10. However, I agree with Bynre's decision on most appropriate parameter F2→F9 (Higher role conflict increases external locus of control). The causal effect for this parameter is most intuitive –role conflict is a type of external factors that have negative influence. Considering its MI (41.343) is so close to the highest seen in this model (43.412), substantive meaningfulness, in my opinion, should have the final say when statistics disagrees moderately. As such, I will continue with setting F2 → F9 free to estimate in model 4. See next section.

3.4 Post hos analysis (Model 4)

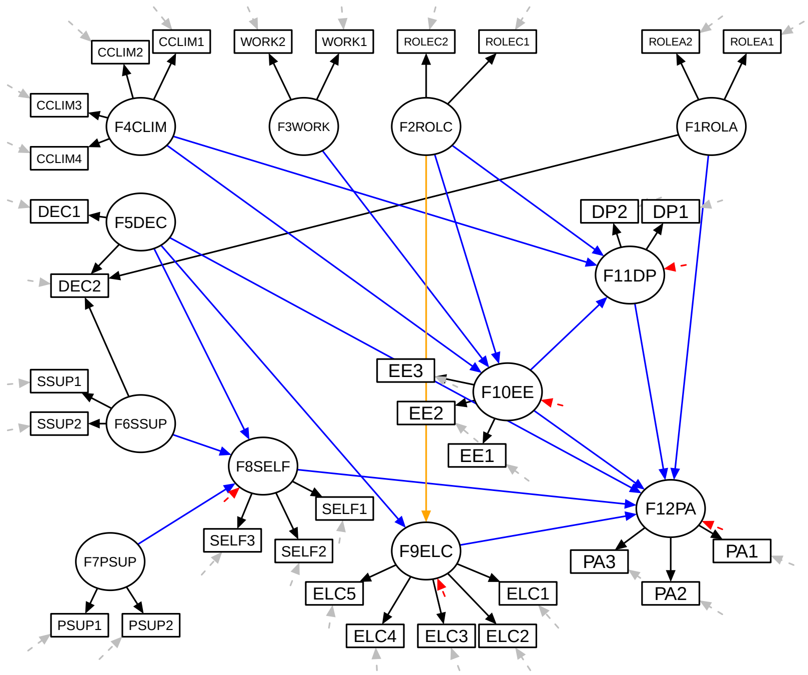
3.4.1 Compare SEM model 4 with preceding models

I defined model 4 as per the conclusion from last section: on the basis of model 3, I set the regression path leading from F2 to F9 (Higher role conflict increases external locus of control) free to estimate (Fig 8). The model fit indices and their comparison with the preceding models were tabulated in table 11.

```
#add new freely estimated parameter to the preceding model
model4 <- paste(model3, "\nF9ELC ~ F2ROLC")

# Estimate the model with the robust (MLM) estimator:
sem4 <- sem(model4, data = mbi, estimator = "MLM", mimic = "Mplus")
```

```
full.sem.diagram(model4, 17, "8", "model4")
```

Figure 8 Modified model (model4) of teacher burnout



Notes: Orange arrow is the regression path set free to estimate in the current model

```
sem.measure4 <- fitMeasures(sem4,    #obtain specified measured.
                            c("chisq.scaled",
                              "df.scaled",
                              "pvalue.scaled",
                              "cfi.scaled",
                              "tli.scaled",
                              "rmsea.scaled",
                              "srmr_bentler",
                              "chisq.scaling.factor")) |>
  t() |>
  round(3)
#combine the 2 sets of indices and save into a new object
sem.compare4 <- rbind(sem.measure1, sem.measure2, sem.measure3, sem.measure4) |>
  data.frame()
sem.compare4.tab <- sem.compare4

#add column "chi square" and place values into it
sem.compare4.tab$"ΔChi-square" <- rep(NA, 4)
sem.compare4.tab$"ΔChi-square"[2] <- chi.diff(sem1, sem2)#calculate chisq difference
sem.compare4.tab$"ΔChi-square"[3] <- chi.diff(sem2, sem3)
sem.compare4.tab$"ΔChi-square"[4] <- chi.diff(sem3, sem4)

#turn named vector to data frame
sem.compare4.tab<- sem.compare4.tab %>%
  select("Chi-square" = chisq.scaled, "DF"=df.scaled,
```

Table 11: Comparison of new (model 4) and preceding models

	Chi-square	DF	p value	ΔChi-square	CFI	TLI	RMSEA	SRMR	CSF
model1	1554.942	427	0	–	0.945	0.936	0.043	0.051	1.117
model2*	1450.985	426	0	91.67	0.950	0.941	0.041	0.046	1.117
Model3†	1406.517	425	0	46.866	0.952	0.944	0.040	0.044	1.117
Model4‡	1366.129	424	0	52.949	0.954	0.946	0.039	0.041	1.118

[*] Model1 + parameter 'F4→F11' set free to estimate

[†] Model2 + parameter 'F5→F12' set free to estimate

[‡] Model3 + parameter 'F2→F9' set free to estimate

```
         "p value"=pvalue.scaled, "ΔChi-square", "CFI"= cfi.scaled,
         "TLI"= tli.scaled, "RMSEA"= rmsea.scaled,
         "SRMR"=srmr_bentler, "CSF"=chisq.scaling.factor)

#There is no chisq-diff value for the first model, so place a "--" in the cell
sem.compare4.tab[1,4] <- "--"

#add row names
rownames(sem.compare4.tab) <- c("model1", "model2*", "Model3†", "Model4‡")

#table aesthetics finetune and print table
sem.compare4.tab |>
  kable(booktab =T,
        #format = "markdown",
        caption = "Comparison of new  (model 4) and preceding models",
        align = "r") |>
  kable_styling() |>
  footnote(symbol = c("Model1 +  parameter 'F4→F11' set free to estimate",
                      "Model2 +  parameter 'F5→F12' set free to estimate",
                      "Model3 +  parameter 'F2→F9' set free to estimate")
           )
```

3.4.2 Examine the parameters of concern for model 4

```
concern.table(sem4, 17, "model4")
```

3.4.3 Comment on the result (model 4)

The estimation of Model 4 yielded a MLM 2 value of 1366.129 (scaling correction factor = 1.118) with 424 degrees of freedom. Values related to the CFI, TLI, RMSEA, and SRMR were .954, .946, 0.039, and 0.041, respectively. Again, the difference in fit between this model (Model 4) and its predecessor (Model 3) was statistically significant (MLM 2[1] = 52.949). See table 11.

As expected, the newly specified parameter (F9 on F2) was found to be statistically significant (Estimate = 0.189; Z-statistics = 7.157). However, once again the three paths leading from F2, F3, and F4 to F10, and from F10 to F12 were all found to be non-significant. Finally, once again, the negative residual associated with Factor 10 appeared.

Table 12: Residual variance of structural regression path and select factors for model4

Parameter*	B†	Beta‡	SE	Z	p-value
Strutural regression path					
F5DEC→F8SELF	0.566	1.177	0.067	8.412	0
F6SSUP→F8SELF	-0.205	-0.648	0.033	-6.147	0
F7PSUP→F8SELF	-0.1	-0.228	0.036	-2.784	0.005
F5DEC→F9ELC	-0.125	-0.208	0.027	-4.638	0
F2ROLC→F10EE	-8.853	-5.875	6.316	-1.402	0.161
F3WORK→F10EE	8.26	6.453	5.359	1.541	0.123
F4CLIM→F10EE	-0.777	-0.226	0.659	-1.179	0.239
F2ROLC→F11DP	0.176	0.138	0.047	3.740	0
F10EE→F11DP	0.3	0.354	0.032	9.537	0
F1ROLA→F12PA	0.271	0.232	0.075	3.626	0
F8SELF→F12PA	0.494	0.225	0.087	5.714	0
F9ELC→F12PA	-0.166	-0.095	0.052	-3.178	0.001
F10EE→F12PA	0.013	0.020	0.027	0.478	0.633
F11DP→F12PA	-0.202	-0.262	0.030	-6.671	0
F4CLIM→F11DP	-0.958	-0.329	0.100	-9.575	0
F5DEC→F12PA	0.432	0.410	0.068	6.387	0
F2ROLC→F9ELC	0.189	0.336	0.026	7.157	0
Selected factors					
F8SELF	0.079	0.682	0.007	10.672	0
F9ELC	0.143	0.774	0.011	13.432	0
F10EE	-0.432	-0.332	0.770	-0.562	0.574348881938236
F11DP	0.605	0.658	0.047	12.880	0
F12PA	0.383	0.695	0.023	16.815	0

Note:

Values highlighted in red should be taken note of

* → indicates regression path

† Crude estimates

‡ Standardized estimates

3.4.4 Model mis-specification

```
#extract needed variables
MI.model4 <- modindices(sem4,
                        standardized = TRUE,
                        sort. = TRUE,
                        maximum.number = 50) |>
  filter(op == "~") |>
  filter(lhs %in%  #When these factors are predicted variables, it is related
           c("F8SELF",  # to the topic of this study
             "F9ELC",
             "F10EE",
             "F11DP",
             "F12PA"))

#adapt to publication style
MI.model4 <- MI.model4 |>
  mutate(Parameter = paste(rhs, "→", lhs)) |>
  select(Parameter, MI = mi, EPC = epc, "std EPC" = sepc.all) |>
  filter(MI>10) |>
  head(3) |>
  mutate("Logics" = c("Sensible and meaningful",
                      "Sensible but not meaningful for the study",
                      "Sensible and not meaningful for the study"))


#add footnote symbol to parameters
  for (i in 1:nrow(MI.model4)){
  symbol <- c("*", "†", "‡")
  Parameter <- unlist(MI.model4$Parameter)
  MI.model4$Parameter[i] <-  paste0(Parameter[i], symbol[i])
  }

#Visualize the table
MI.model4 |>
  kable(digits = 3,
        booktab = T,
        linesep = "",
        caption = "Selected modification indices for initial model") |>
  kable_styling(latex_options = "striped") |>
  row_spec(1, color = "red") |>
  footnote(general =
             "Parameter highlighted in red is selected for modification",
    symbol = c("Lower self-esteem causes severer external locus of control",
             "Higher personal accomplishment increases self-esteem",
             "Higher personal accomplishment decreases external locus of control"))
```

The parameter F8→F9 shows up again but this time it has the highest MI among its peers. See tab 13. In the preceding model (model 3) I did not adopt it only because its relatively lower MI albeit good sensibility and meaningfulness. Given that other two competitor parameters in the current model were bad in meaningfulness (because they regress risk factors on the outcome of interest, but we want the opposite), we could decide on adopting F8→F9 (Lower self-esteem causes severer external locus of control) without much ado.

Table 13: Selected modification indices for initial model

Parameter	MI	EPC	std EPC	Logics
F8SELF → F9ELC*	38.685	-0.261	-0.208	Sensible and meaningful
F12PA → F8SELF†	37.407	0.302	0.662	Sensible but not meaningful for the study
F12PA → F9ELC‡	30.392	-0.265	-0.464	Sensible and not meaningful for the study

Note:

Parameter highlighted in red is selected for modification

* Lower self-esteem causes severer external locus of control

† Higher personal accomplishment increases self-esteem

‡ Higher personal accomplishment decreases external locus of control

3.5 Post hos analysis (Model 5)

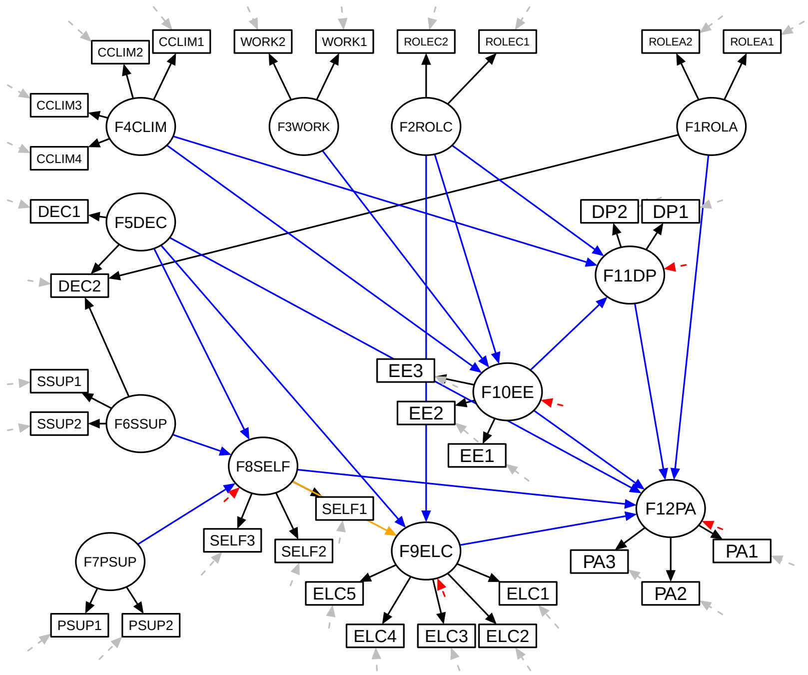
3.5.1 Compare SEM model 5 with preceding models

I defined model 5 as per the conclusion from last section: on the basis of model 4, I set the regression path leading from F8 to F9 (Lower self-esteem causes severer external locus of control) free to estimate (Fig 9). The model fit indices and their comparison with the preceding models were tabulated in table 14.

```
#add new freely estimated parameter to the preceding model
model5 <- paste(model4, "\nF9ELC ~ F8SELF")

# Estimate the model with the robust (MLM) estimator:
sem5 <- sem(model5, data = mbi, estimator = "MLM", mimic = "Mplus")
```

```
full.sem.diagram(model5, 18, "9", "model5")
```

Figure 9 Modified model (model5) of teacher burnout



Notes: Orange arrow is the regression path set free to estimate in the current model

```
sem.measure5 <- fitMeasures(sem5,    #obtain specified measured.
                            c("chisq.scaled",
                              "df.scaled",
                              "pvalue.scaled",
                              "cfi.scaled",
                              "tli.scaled",
                              "rmsea.scaled",
                              "srmr_bentler",
                              "chisq.scaling.factor")) |>
  t() |>
  round(3)
#combine the 2 sets of indices and save into a new object
sem.compare5 <- rbind(sem.measure1, sem.measure2,
                      sem.measure3, sem.measure4, sem.measure5) |>
  data.frame()
sem.compare5.tab <- sem.compare5

#add column "chi square" and place values into it
sem.compare5.tab$"ΔChi-square" <- rep(NA, 5)
sem.compare5.tab$"ΔChi-square"[2] <- chi.diff(sem1, sem2)#calculate chisq difference
sem.compare5.tab$"ΔChi-square"[3] <- chi.diff(sem2, sem3)
sem.compare5.tab$"ΔChi-square"[4] <- chi.diff(sem3, sem4)
sem.compare5.tab$"ΔChi-square"[5] <- chi.diff(sem4, sem5)

#turn named vector to data frame
```

Table 14: Comparison of new (model 5) and preceding models

	Chi-square	DF	p value	ΔChi-square	CFI	TLI	RMSEA	SRMR	CSF
model1	1554.942	427	0	–	0.945	0.936	0.043	0.051	1.117
model2*	1450.985	426	0	91.67	0.950	0.941	0.041	0.046	1.117
Model3†	1406.517	425	0	46.866	0.952	0.944	0.040	0.044	1.117
Model4‡	1366.129	424	0	52.949	0.954	0.946	0.039	0.041	1.118
Model5§	1331.636	423	0	29.877	0.955	0.948	0.039	0.039	1.117

* Model1 + parameter 'F4→F11' set free to estimate

† Model2 + parameter 'F5→F12' set free to estimate

‡ Model3 + parameter 'F2→F9' set free to estimate

§ Model5 + parameter 'F8→F9' set free to estimate

```
sem.compare5.tab<- sem.compare5.tab %>%
  select("Chi-square" = chisq.scaled, "DF"=df.scaled,
         "p value"=pvalue.scaled, "ΔChi-square", "CFI"= cfi.scaled,
         "TLI"= tli.scaled, "RMSEA"= rmsea.scaled,
         "SRMR"=srmr_bentler, "CSF"=chisq.scaling.factor)

#There is no chisq-diff value for the first model, so place a "--" in the cell
sem.compare5.tab[1,4] <- "--"

#add row names
rownames(sem.compare5.tab) <- c("model1", "model2*",
                                "Model3†", "Model4‡",
                                "Model5§")

#table aesthetics fine-tune and print table
sem.compare5.tab |>
  kable(booktab =T,
        #format = "markdown",
        caption = "Comparison of new  (model 5) and preceding models",
        align = "r") |>
  kable_styling() |>
  footnote(symbol = c("Model1 +  parameter 'F4→F11' set free to estimate",
                      "Model2 +  parameter 'F5→F12' set free to estimate",
                      "Model3 +  parameter 'F2→F9' set free to estimate",
                      "Model5 +  parameter 'F8→F9' set free to estimate")
           )
```

3.5.2 Examine the parameters of concern for model 5

```
concern.table(sem5, 18, "model5")
```

3.5.3 Comment on the result (model 5)

The estimation of Model 5 yielded a MLM 2(423) value of 1331.636 (scaling correction factor = 1.117), with other fit indices as follows: CFI = 0.955, TLI = 0.948, RMSEA = 0.039, and SRMR = 0.039. Again, the

Table 15: Residual variance of structural regression path and select factors for model5

Parameter*	B†	Beta‡	SE	Z	p-value
Strutural regression path					
F5DEC→F8SELF	0.591	1.221	0.074	8.038	0
F6SSUP→F8SELF	-0.22	-0.696	0.037	-5.960	0
F7PSUP→F8SELF	-0.109	-0.249	0.039	-2.826	0.005
F5DEC→F9ELC	-0.076	-0.125	0.027	-2.803	0.005
F2ROLC→F10EE	-6.508	-4.323	3.440	-1.892	0.058
F3WORK→F10EE	6.262	4.907	2.914	2.149	0.032
F4CLIM→F10EE	-0.718	-0.209	0.492	-1.460	0.144
F2ROLC→F11DP	0.173	0.135	0.047	3.688	0
F10EE→F11DP	0.302	0.356	0.031	9.640	0
F1ROLA→F12PA	0.259	0.222	0.074	3.512	0
F8SELF→F12PA	0.521	0.237	0.090	5.815	0
F9ELC→F12PA	-0.162	-0.092	0.055	-2.957	0.003
F10EE→F12PA	0.013	0.019	0.027	0.463	0.643
F11DP→F12PA	-0.202	-0.262	0.030	-6.680	0
F4CLIM→F11DP	-0.959	-0.329	0.100	-9.579	0
F5DEC→F12PA	0.417	0.392	0.065	6.386	0
F2ROLC→F9ELC	0.174	0.310	0.026	6.744	0
F8SELF→F9ELC	-0.257	-0.205	0.048	-5.337	0
Selected factors					
F8SELF	0.079	0.682	0.007	10.672	0
F9ELC	0.143	0.774	0.011	13.432	0
F10EE	-0.432	-0.332	0.770	-0.562	0.574
F11DP	0.605	0.658	0.047	12.880	0
F12PA	0.383	0.695	0.023	16.815	0

Note:

Values highlighted in red should be taken note of

* → indicates regression path

† Crude estimates

‡ Standardized estimates

difference in fit between this model (Model 5) and its predecessor (Model 4) was found to be statistically significant (MLM 2[1] = 29.877). See table 14.

Once again, the newly specified parameter (F9 on F8) was found to be statistically significant and accompanied by the correct sign (Estimate = -0.257; Z-statistics = -5.337). However, as with Model 4, again two of the three paths leading to F10 (F2 → F10; F4 → F10), and one leading from F10 to F12 remained statistically non-significant. Besides, the negative residual associated with Factor 10 remained unchanged. See tab 15.

3.5.4 Model mis-specification

```
#extract needed variables
MI.model5 <- modindices(sem5,
                  standardized = TRUE,
                  sort. = TRUE,
                  maximum.number = 50) |>
  filter(op == "~") |>
  filter(lhs %in%  #When these factors are predicted variables, it is related
           c("F8SELF",  # to the topic of this study
             "F9ELC",
             "F10EE",
             "F11DP",
             "F12PA"))

#adapt to publication style
MI.model5 <- MI.model5 |>
  mutate(Parameter = paste(rhs, "→", lhs)) |>
  select(Parameter, MI = mi, EPC = epc, "std EPC" = sepc.all) |>
  head(2) |>
  mutate("Logics" = c("Sensible and meaningful",
                      "Sensible but not meaningful for the study"))

#add footnote symbol to parameters
  for (i in 1:nrow(MI.model5)){
  symbol <- c("*", "†", "‡")
  Parameter <- unlist(MI.model5$Parameter)
  MI.model5$Parameter[i] <-  paste0(Parameter[i], symbol[i])
  }

#Visualize the table
MI.model5 |>
  kable(digits = 3,
        booktab = T,
        linesep = "",
        caption = "Selected modification indices for initial model") |>
  kable_styling(latex_options = "striped") |>
  row_spec(1, color = "red") |>
  footnote(general =
             "Parameter highlighted in red is selected for modification",
    symbol = c("Lower self-esteem causes severer emotional exhaustion",
             "Higher depersonalization causes severer role conflict"))
```

This output reveals the structural path leading from Self-Esteem to Emotional Exhaustion (F8 → F10)

Table 16: Selected modification indices for initial model

Parameter	MI	EPC	std EPC	Logics
F8SELF → F10EE*	30.096	-1.068	-0.318	Sensible and meaningful
F11DP → F8SELF†	25.671	-0.066	-0.188	Sensible but not meaningful for the study

Note:

Parameter highlighted in red is selected for modification

* Lower self-esteem causes severer emotional exhaustion

† Higher depersonalization causes severer role conflict

as having the largest MI value. Given that the fact that it seems reasonable that teachers who exhibit high levels of self-esteem may exhibit low levels of emotional exhaustion, the model was re-estimated once again, with this path freely estimated. See tab .One other interesting reason to do so, cited from Byrne is that *"Because Factor 10 has been problematic regarding the estimation of its residual in yielding a negative variance, it would seem likely that if this parameter were to be included in the model, this undesirable result may finally be resolved."*

3.6 Post hos analysis (Model 6)

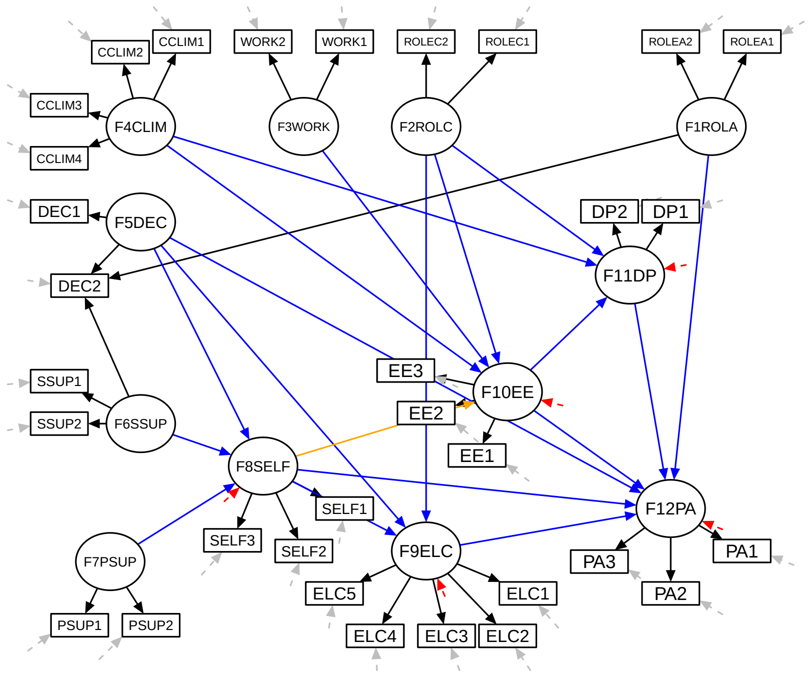
3.6.1 Compare SEM model 6 with preceding models

I defined model 6 as per the conclusion from last section: on the basis of model 5, I set the regression path leading from F8 to F9 (Lower self-esteem causes severer emotional exhaustion) free to estimate (Fig 10). The model fit indices and their comparison with the preceding models were tabulated in table 16.

```
#add new freely estimated parameter to the preceding model
model6 <- paste(model5, "\nF10EE ~ F8SELF")

# Estimate the model with the robust (MLM) estimator:
sem6 <- sem(model6, data = mbi, estimator = "MLM", mimic = "Mplus")
```

```
full.sem.diagram(model6, 19, "9", "model6")
```

Figure 9 Modified model (model6) of teacher burnout



Notes: Orange arrow is the regression path set free to estimate in the current model

```
sem.measure6 <- fitMeasures(sem6,    #obtain specified measured.
                            c("chisq.scaled",
                              "df.scaled",
                              "pvalue.scaled",
                              "cfi.scaled",
                              "tli.scaled",
                              "rmsea.scaled",
                              "srmr_bentler",
                              "chisq.scaling.factor")) |>
  t() |>
  round(3)
#combine the 2 sets of indices and save into a new object
sem.compare6 <- rbind(sem.measure1, sem.measure2,
                      sem.measure3, sem.measure4,
                      sem.measure5, sem.measure6) |>
  data.frame()
sem.compare6.tab <- sem.compare6

#add column "chi square" and place values into it
sem.compare6.tab$"ΔChi-square" <- rep(NA, 6)
sem.compare6.tab$"ΔChi-square"[2] <- chi.diff(sem1, sem2)#calculate chisq difference
sem.compare6.tab$"ΔChi-square"[3] <- chi.diff(sem2, sem3)
sem.compare6.tab$"ΔChi-square"[4] <- chi.diff(sem3, sem4)
sem.compare6.tab$"ΔChi-square"[5] <- chi.diff(sem4, sem5)
sem.compare6.tab$"ΔChi-square"[6] <- chi.diff(sem5, sem6)
```

Table 17: Comparison of new (model 6) and preceding models

	Chi-square	DF	p value	ΔChi-square	CFI	TLI	RMSEA	SRMR	CSF
model1	1554.942	427	0	–	0.945	0.936	0.043	0.051	1.117
model2*	1450.985	426	0	91.67	0.950	0.941	0.041	0.046	1.117
Model3†	1406.517	425	0	46.866	0.952	0.944	0.040	0.044	1.117
Model4‡	1366.129	424	0	52.949	0.954	0.946	0.039	0.041	1.118
Model5§	1331.636	423	0	29.877	0.955	0.948	0.039	0.039	1.117
Model6	1297.489	422	0	23.358	0.957	0.949	0.038	0.039	1.116

[*] Model1 + parameter 'F4→F11' set free to estimate
[†] Model2 + parameter 'F5→F12' set free to estimate
[‡] Model3 + parameter 'F2→F9' set free to estimate
[§] Model5 + parameter 'F8→F9' set free to estimate
[¶] Model6 + parameter 'F8→F10' set free to estimate

```r
#turn named vector to data frame
sem.compare6.tab<- sem.compare6.tab %>%
  select("Chi-square" = chisq.scaled, "DF"=df.scaled,
         "p value"=pvalue.scaled, "ΔChi-square", "CFI"= cfi.scaled,
         "TLI"= tli.scaled, "RMSEA"= rmsea.scaled,
         "SRMR"=srmr_bentler, "CSF"=chisq.scaling.factor)

#There is no chisq-diff value for the first model, so place a "--" in the cell
sem.compare6.tab[1,4] <- "--"

#add row names
rownames(sem.compare6.tab) <- c("model1", "model2*",
                                "Model3†", "Model4‡",
                                "Model5§", "Model6")

#table aesthetics fine-tune and print table
sem.compare6.tab |>
  kable(booktab =T,
        #format = "markdown",
        caption = "Comparison of new  (model 6) and preceding models",
        align = "r") |>
  kable_styling() |>
  footnote(symbol = c("Model1 +  parameter 'F4→F11' set free to estimate",
                      "Model2 +  parameter 'F5→F12' set free to estimate",
                      "Model3 +  parameter 'F2→F9' set free to estimate",
                      "Model5 +  parameter 'F8→F9' set free to estimate",
                      "Model6 +  parameter 'F8→F10' set free to estimate")
           )
```

3.6.2 ### Examine the parameters of concern for model 6

3.6.3 Comment on the result (model 6)

3.6.4 Model mis-specification

Note: Here, we will use the sem() function for the estimation, instead of the cfa() function, as we are now working with a full SEM (i.e., CFA + regression paths).

3.7 Exercise 4.2

Proceed **step by step** following the guidelines given in the lecture material, i.e., implement the modifications **one at a time**, testing and studying each step. See (and report) how the fit improves and which parameters are suggested to be modified. Please be careful! There will (always) be a lot of suggestions… Do not list all the MIs (only a few of them are useful!), try to keep your report as concise as possible.

Note: A good way to proceed is to collect the necessary information (i.e., which parameter was modified and how, MI, EPC, chi-square, df, CFI, TLI, scaling correction factor, RMSEA, and SRMR) of each modelling step to a **table** (in a way or another). (Some examples in R code were given in Assignment 3, consult also the reports by other students, if you do not know how to proceed.) **Such tables makes it easy to see how the results of the modelling develop through each step.**

The best practice is to build the tables step by step: In the first table you will have only one row, then two rows, then three rows etc., and in the final version of the table you will have all the steps collected together on k rows, representing the k steps of the modelling process.

3.7.1 Calculating the MLM χ^2 difference tests

Calculate the MLM χ^2 difference tests between the consecutive models of the above steps, as advised in the lecture material (p.14-15). Do those calculations in detail at least once or twice so that you get the idea.

Note: The formulas are simpler than they are in Byrne's book (p.168-169), where both MLM and ML estimations are needed. For more information, see: https://statmodel.com/chidiff.shtml.

For the calculations, you may use R (of course!) or Excel, or some ready-made calculation forms found on the web, such as https://www.thestatisticalmind.com/calculators/SBChiSquareDifferenceTest.htm.

```
# (copy and modify the R codes given earlier)
```

3.8 Exercise 4.3

Draw the graph of the final model and present its fit indices and the essential, standardized parameter estimates. **Pay attention to the factor correlations.**

Compare the initial and final graphs and make sure that you understand the whole modelling process and the final conclusions.

```
# (copy and modify the R codes given earlier)
```