

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

Rong Guang

2023-02-10

Contents

1 SEM & teacher burnout	1
1.1 Exercise 4.1	1
2 Preparation	1
2.1 Read in the data set	1
2.2 Write functions	1
2.3 Inspect the data	2
3 Testing the for the validity of causal structure of burnout	10
3.1 Initial full structural equation model (hypothesized model modified according to CFA)	10
3.2 Post hos analysis (Model 2)	18
3.3 Post hos analysis (Model 3)	25
3.4 Post hos analysis (Model 4)	27
3.5 Post hos analysis (Model 5)	27
3.6 Post hos analysis (Model 6)	27
3.7 Exercise 4.2	28
3.8 Exercise 4.3	28

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 aesthetics 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

2.3 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

			Central tendency		Dispersion tendency			
	n	n of NA	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

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

	n	n of NA	Central tendency		Dispersion tendency			
			Mean	Median	SD	Min	Max	Q1~Q3
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								
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

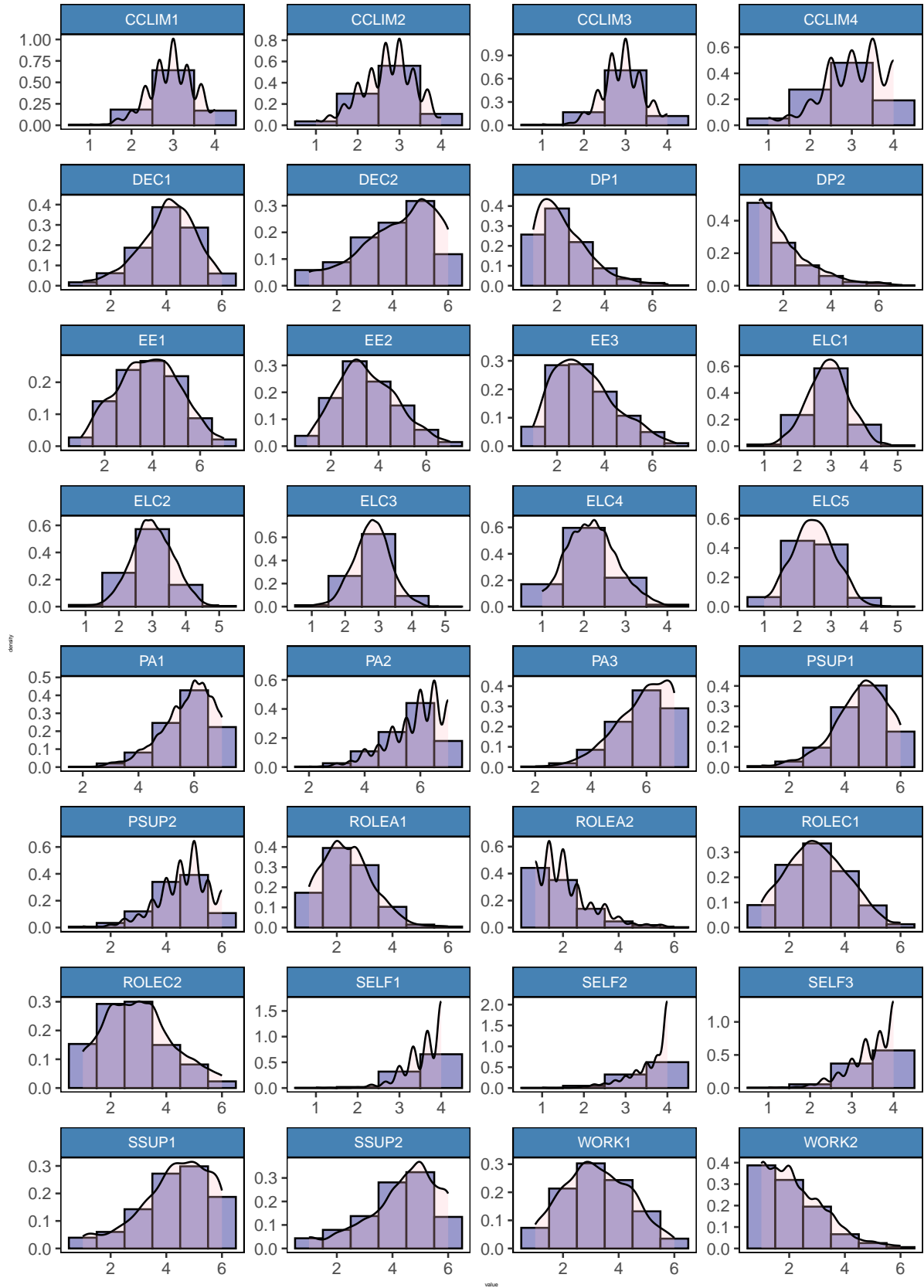
2.3.3 Visualization

(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(ggribes)
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

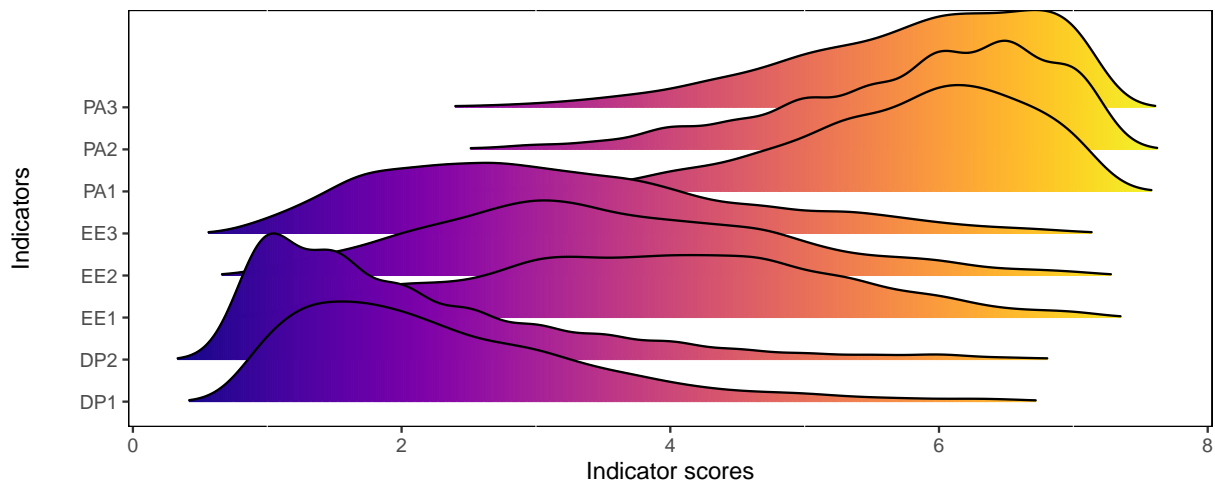
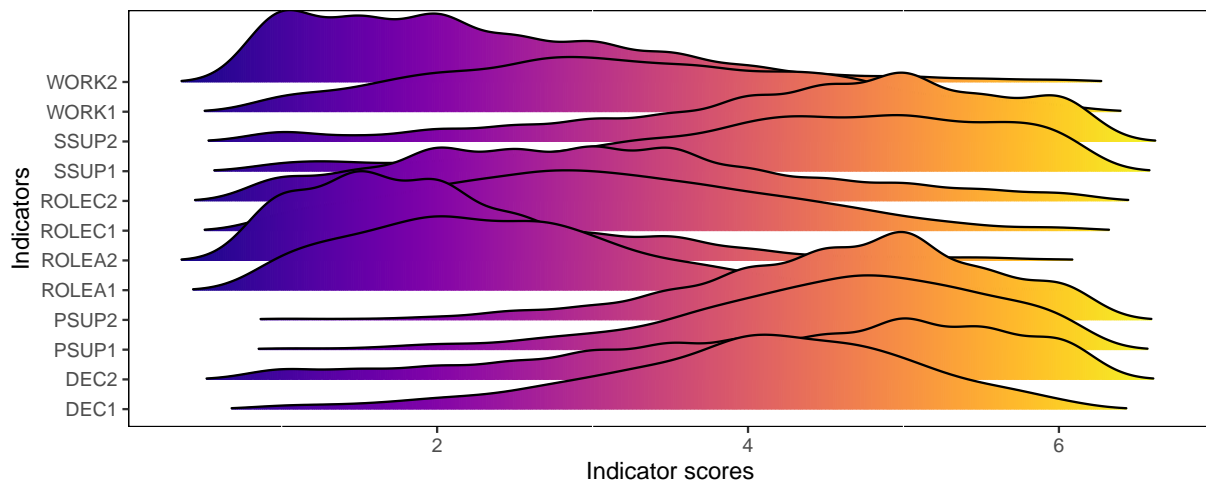


Fig2 (b). Distribution of indicator scores from TSS instrument



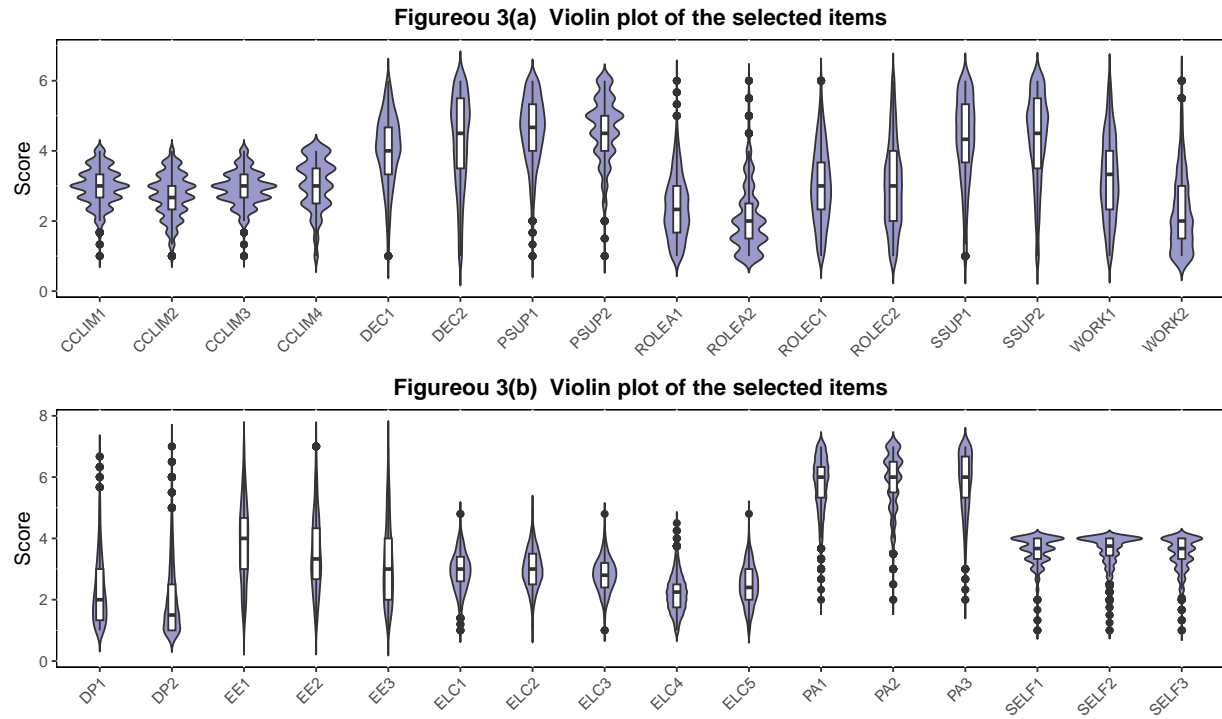
Clearly, within each instrument, indicators for same factor tend to show similar distribution features.

(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)")

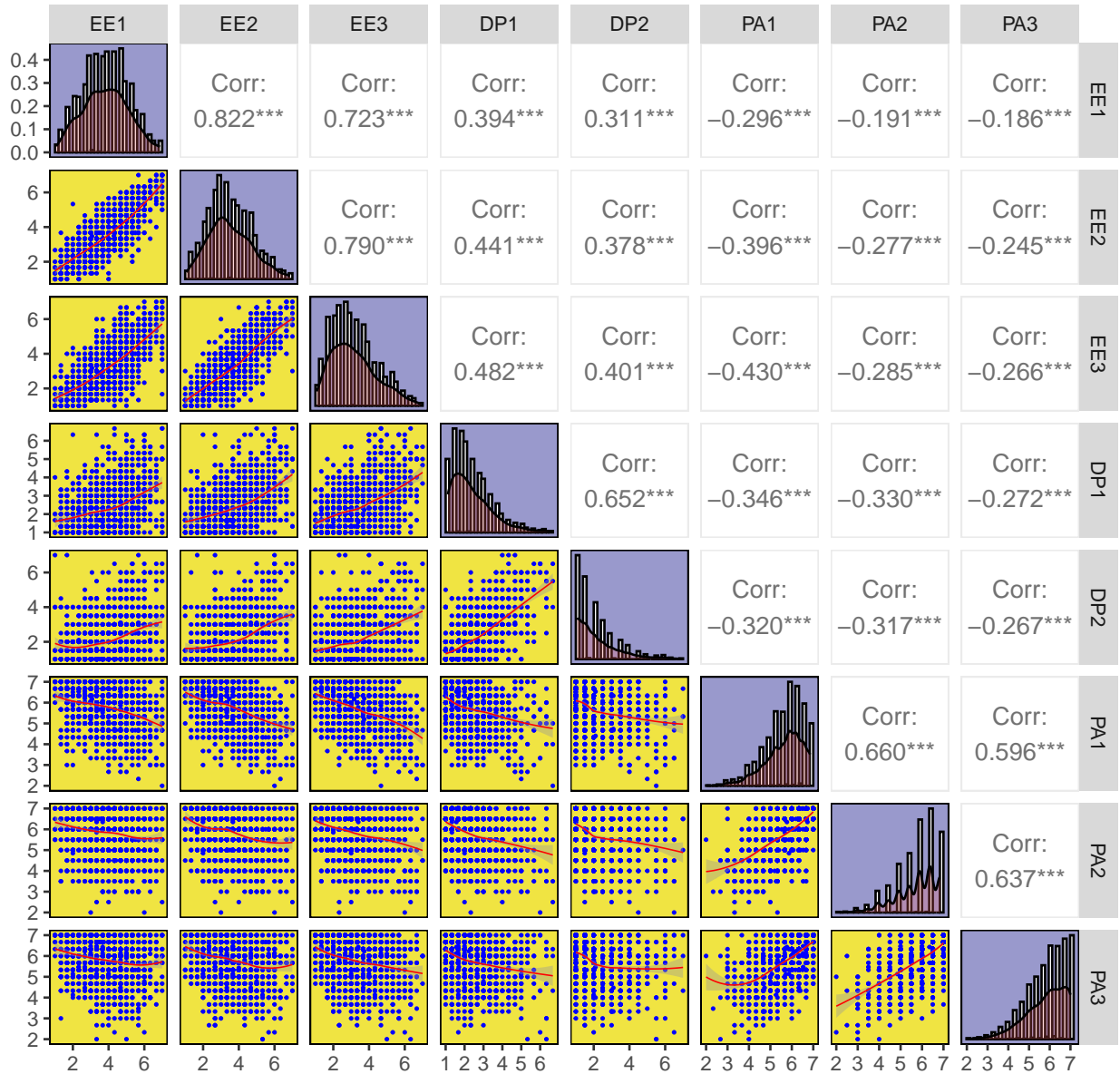
b <- mbi |>
  select(17:32) |>
  violin.box(fig.num = "3(b)")
library(patchwork)
a/b
```

(3) Correlation among items

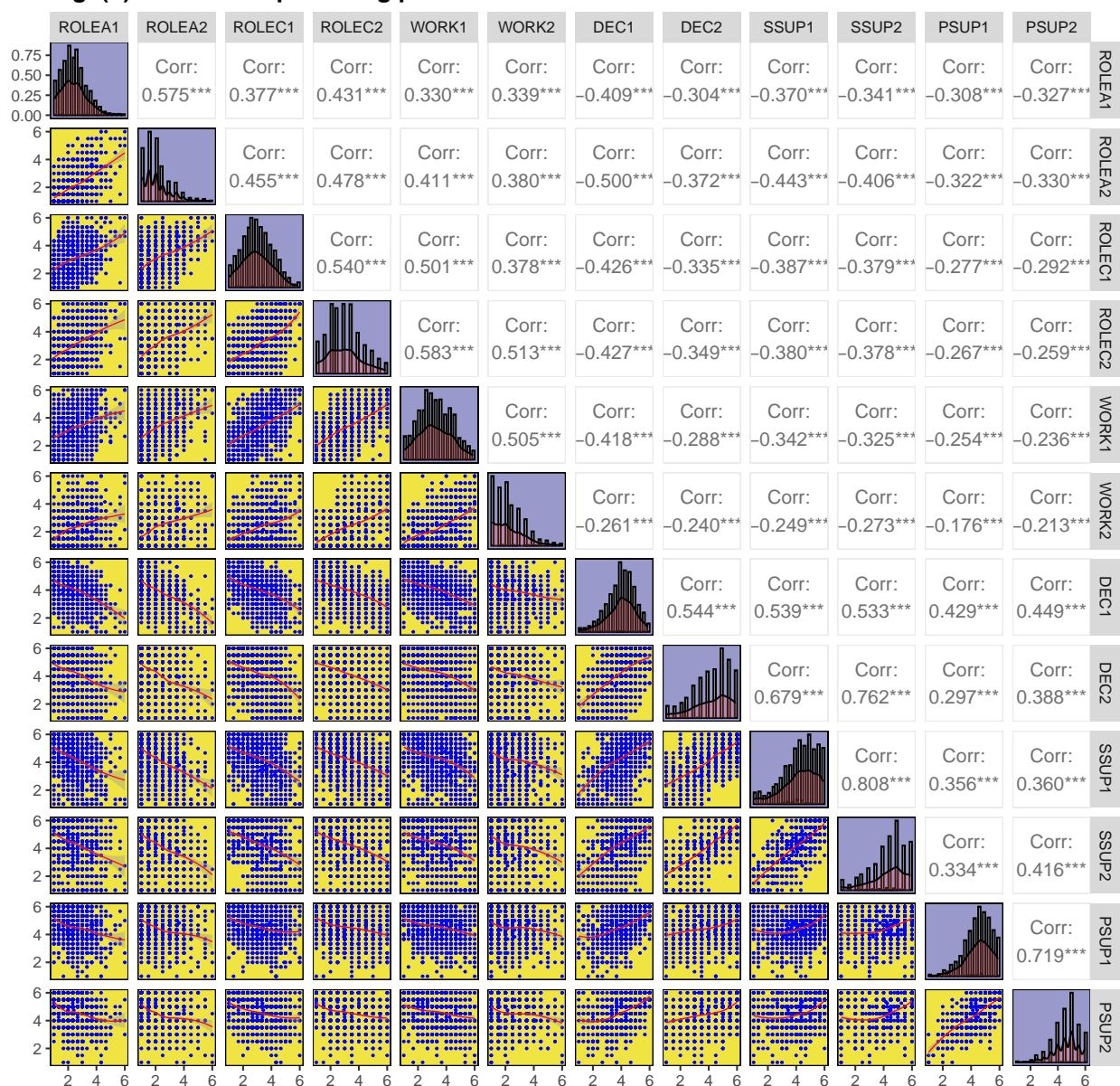
```
#draw it
mbi |> select(starts_with("EE")|starts_with("DP")|starts_with("PA")) |>
  ggpairs(lower =
    list(continuous = my.fun.smooth), #lower half show points with fitted line
    diag =
    list(continuous = my.fun.density), #diagonal grids show density plots
    title = "Fig4(a). Relationships 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"
```

```

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

#define the positions 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.

```

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,

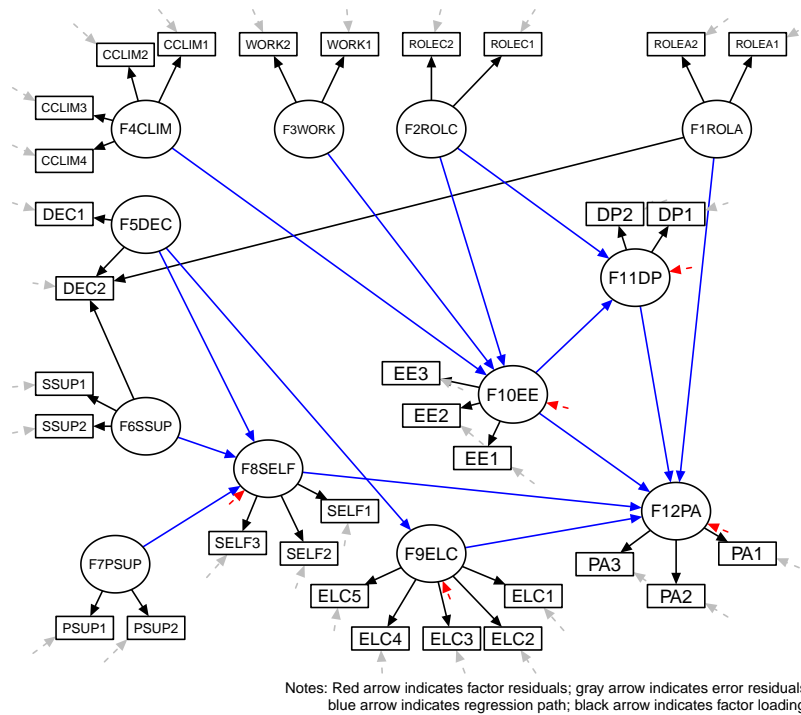
```

```

    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



3.1.2 Estimate the SEM model (initial)

```

library(lavaan)
modell1 <- initial_model # defined above

# Estimate the model with the robust (MLM) estimator:
sem1 <- sem(modell1, 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 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)

options(scipen = 999)
#regression path estimates
sem.parameter <- parameterEstimates(sem1, standardized=TRUE) |> # obtain estimates
  filter(op == "~") |> #select "is measured by" rows
  select('DV*' = lhs, #left hand side column
        'IV*' = rhs, #right hand side column
        'B†' = est, #estimates
        '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")
                                     )

#furhter aesthetics
sem.parameter |>
  kable(digits = 3, #rounded to 3
        #format="latex", #Latex markdown
        booktabs=TRUE, #Latex booktabs
        linesep = "",
        caption= "Structural regression path and residual variance estimates.",
        escape = F) %>% #caption
  kable_styling(latex_options = "striped") %>% #gray every other row
  footnote(general = "Rows in bold have insignificant parameters.",
          symbol = c("DV: dependent variable; IV: independent variable",
                    "Crude estimates",
                    "Standardized estimates"))
```

Table 3: Structural regression path and residual variance estimates.

DV*	IV*	B†	Beta‡	SE	Z	p-value
F8SELF	F5DEC	0.475	0.990	0.050	9.462	0
F8SELF	F6SSUP	-0.155	-0.490	0.025	-6.194	0
F8SELF	F7PSUP	-0.066	-0.150	0.029	-2.307	0.021
F9ELC	F5DEC	-0.288	-0.476	0.021	-13.784	0
F10EE	F2ROLC	-8.707	-5.765	6.310	-1.380	0.168
F10EE	F3WORK	8.082	6.325	5.321	1.519	0.129
F10EE	F4CLIM	-0.930	-0.272	0.658	-1.414	0.157
F11DP	F2ROLC	0.258	0.203	0.048	5.356	0
F11DP	F10EE	0.373	0.444	0.033	11.377	0
F12PA	F1ROLA	-0.071	-0.062	0.046	-1.548	0.122
F12PA	F8SELF	0.472	0.217	0.085	5.562	0
F12PA	F9ELC	-0.208	-0.121	0.049	-4.264	0
F12PA	F10EE	-0.064	-0.098	0.025	-2.564	0.01
F12PA	F11DP	-0.218	-0.281	0.031	-7.115	0

Note:

Rows in bold have insignificant parameters.

* DV: dependent variable; IV: independent variable

† Crude estimates

‡ Standardized estimates

```
#Variance
type <- "Total variance" #create a new row clarifying types of variance
#write a function for minus calculation
minus <- function(x,y) {x - y}

variance <- parameterEstimates(sem1, standardized=TRUE) |> #obtain estimates
  filter(op == "~~") #select "is correlated with" rows
variance <- variance[minus(sum(32,12), 5):sum(32,12),] #subset needed rows (variance row)
variance <- cbind(type, variance) #add column
sem.tab.variance <- variance %>%select(Type = type, #select and rename variables
  Factor=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,
```


Table 4: Residual variance of the dependent variables for initial model

Type	Factor	B*	Beta†	SE	Z	p-value
Total variance	F7PSUP	0.595	1.000	0.041	14.351	0.000
Total variance	F8SELF	0.079	0.682	0.007	10.672	0.000
Total variance	F9ELC	0.143	0.774	0.011	13.432	0.000
Total variance	F10EE	-0.432	-0.332	0.770	-0.562	0.574
Total variance	F11DP	0.605	0.658	0.047	12.880	0.000
Total variance	F12PA	0.383	0.695	0.023	16.815	0.000

Note:

Negative crude residual variance will be highlighted in red

* Crude estimates

† Standardized estimates

```

                                "red",
                                "black")
                                )

sem.tab.variance |>
  kable(digits = 3, #rounded
        #format="markdown", #Latex markdown
        booktabs=TRUE, #Latex booktabs
        caption=
          "Residual variance of the dependent variables for initial model",
        escape = F) |> #caption
  kable_styling(latex_options = "striped") |> # gray every other row
  footnote(general =
    "Negative crude residual variance will be highlighted in red",
    symbol = c("Crude estimates",
               "Standardized estimates"))

```

3.1.3 Comments on the result (initial model)

Here we see that the rescaled χ^2 value (i.e., the MLM χ^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 χ^2 were multiplied by 1.127, it would approximate the uncorrected ML χ^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 and 4.

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 \rightarrow Emotional Exhaustion), F10 on F3 (Work Overload \rightarrow Emotional Exhaustion), F10 on F4 (Classroom Climate \rightarrow Emotional Exhaustion), and (b) F12 on F1 (Role Ambiguity \rightarrow Personal Accomplishment).

For the negative variance for the residual associated with Factor 10 (highlighted in red in table 4), 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

```
#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) |>
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("∗", "†", "‡")
  Parameter <- unlist(MI.model1$Parameter)
  MI.model1$Parameter[i] <- paste0(Parameter[i], symbol[i])
}

#Visualize the table
MI.model1 |>
kable(digits = 3,
      booktab = T,
      linesep = "",
      caption = "Select 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("Poorer Classroom climate leads to worsening depersonalization",
             "Less workload leads to worsening depersonalization",
             "Higher accomplishment results in increased self-esteem"))
```

Table 5: Select 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

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 indice comparison between the current model (model2) and the initial model was tabulated in table .

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

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

# Numerical summary of the model:
sem.summary.mlm.a(sem2, 12, 32, "mlm", "Model fit indices for model 2") |>
  unlist()
```

Table 6: Model fit indices for model 2

Measure	Value
chi square	1450.985
df	426.000
p value	0.000
CFI	0.950
TLI	0.941
RMSEA	0.041
RMSEA p value	1.000
SRMR	0.046
CSF	1.117

A function for calculating 2 difference was defined.

```
chi.diff <- function(sem1, sem2){
  measure0 <- fitMeasures(sem1,
    c("df.scaled",
      "chisq.scaling.factor",
      "chisq.scaled")) |>
```

```

    as.vector()
measure1 <- fitMeasures(sem2,
                        c("df.scaled",
                          "chisq.scaling.factor",
                          "chisq.scaled")) |>

    as.vector()
cd <-
  (measure0[1]*measure0[2]-measure1[1]*measure1[2])/(measure0[1]-measure1[1])
TRd <-
  (measure0[3]*measure0[2]-measure1[3]*measure1[2])/cd
TRd <- TRd |> round(3)
print(TRd)
}

```

```

#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() |>
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("2", "DF", "p value", "CFI", "TLI",
                           "RMSEA", "SRMR", "CSF")
#turn named vector to data frame
sem.compare2.tab<- sem.compare2 %>%
  mutate("Δ 2" = chi.diff(sem1, sem2)) |>
  select("2", "DF", "p value", "Δ 2", "CFI", "TLI", "RMSEA", "SRMR", "CSF")

```

```
## [1] 91.67
```

Table 7: Comparison of new and preceding models

	2	DF	p value	$\Delta 2$	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

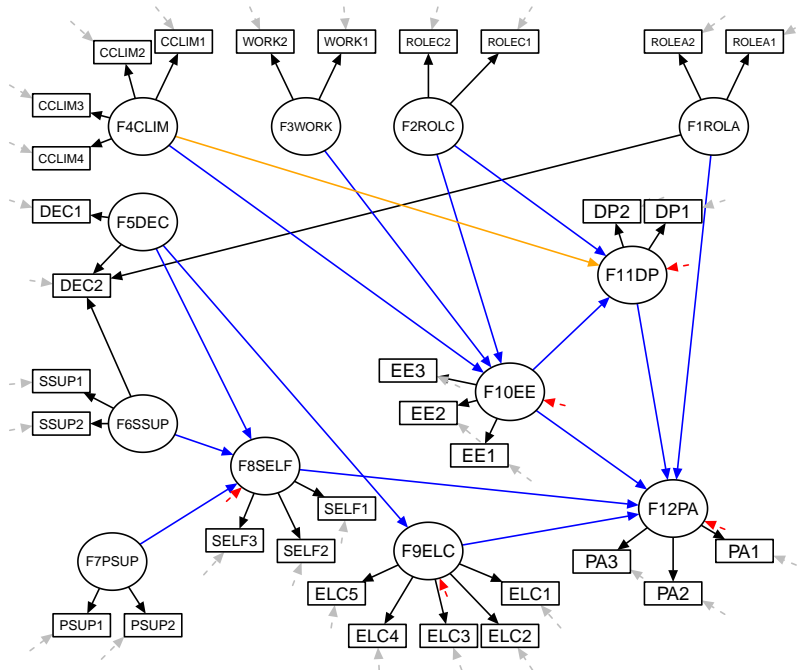
* Model1 + parameter 'F4→F11' set free to estimate

```
sem.compare2.tab[1,4] <- "--"
rownames(sem.compare2.tab) <- c("model1", "model2*")

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")

semPaths(semPlotModel(model2),
         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),
                        "orange",
                        rep("gray",32),
                        rep("red",5)),
         residuals = T,
         layout = m,
         nCharNodes=0,
         optimizeLatRes = T,
         exoVar = F)
title(main = list("Figure 6. Modified model (Model 2) of teacher burnout",
                  cex = 1.5, font = 1),
      outer = F, line = -1)
title(sub =
      "Notes: Orange arrow is the regression path set free to estimate in the current model",
      line = 0, adj = 0.7)
```

Figure 6. Modified model (Model 2) of teacher burnout



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

3.2.2 Estimate SEM model 2

```
#regression path estimates
sem.parameter <- parameterEstimates(sem2, standardized=TRUE) |> # obtain estimates
  filter(op == "~") |> #select "is measured by" rows
  select('DV*' = lhs, #left hand side column
        'IV*' = rhs, #right hand side column
        'B†' = est, #estimates
        '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",
  ),
```

Table 8: Structural regression path and residual variance estimates.

DV*	IV*	B†	Beta‡	SE	Z	p-value
F8SELF	F5DEC	0.474	0.989	0.050	9.453	0
F8SELF	F6SSUP	-0.155	-0.489	0.025	-6.198	0
F8SELF	F7PSUP	-0.066	-0.150	0.029	-2.308	0.021
F9ELC	F5DEC	-0.288	-0.476	0.021	-13.791	0
F10EE	F2ROLC	-8.396	-5.553	5.945	-1.412	0.158
F10EE	F3WORK	7.856	6.151	5.034	1.561	0.119
F10EE	F4CLIM	-0.563	-0.164	0.633	-0.888	0.374
F11DP	F2ROLC	0.173	0.135	0.046	3.726	0
F11DP	F10EE	0.299	0.354	0.031	9.627	0
F12PA	F1ROLA	-0.069	-0.060	0.046	-1.489	0.137
F12PA	F8SELF	0.473	0.217	0.084	5.599	0
F12PA	F9ELC	-0.204	-0.118	0.049	-4.190	0
F12PA	F10EE	-0.058	-0.089	0.024	-2.411	0.016
F12PA	F11DP	-0.228	-0.297	0.030	-7.533	0
F11DP	F4CLIM	-0.969	-0.334	0.100	-9.642	0

Note:

Rows in bold have insignificant parameters.

* DV: dependent variable; IV: independent variable

† Crude estimates

‡ Standardized estimates

```

                                "black")
                                )
#furhter aesthetics
sem.parameter |>
  kable(digits = 3, #rounded to 3
        #format="latex", #Latex markdown
        booktabs=TRUE, #Latex booktabs
        linesep = "",
        caption= "Structural regression path and residual variance estimates.",
        escape = F) %>% #caption
  kable_styling(latex_options = "striped") %>% #gray every other row
  footnote(general = "Rows in bold have insignificant parameters.",
          symbol = c("DV: dependent variable; IV: independent variable",
                    "Crude estimates",
                    "Standardized estimates"))

#Variance
type <- "Total variance" #create a new row clarifying types of variance

variance <- parameterEstimates(sem2, standardized=TRUE) |> #obtain estimates
  filter(op == "~") #select "is correlated with" rows
variance <- variance[minus(sum(32,12), 5):sum(32,12),] #subset needed rows (variance row)
variance <- cbind(type, variance) #add column
sem.tab.variance <- variance %>%select(Type = type, #select and rename variables
  Factor=rhs, #right hand side column
  'B*'='est, #estimates

```

Table 9: Residual variance of the dependent variables for initial model

Type	Factor	B*	Beta†	SE	Z	p-value
Total variance	F7PSUP	0.595	1.000	0.042	14.329	0.0
Total variance	F8SELF	0.079	0.682	0.007	10.678	0.0
Total variance	F9ELC	0.143	0.774	0.011	13.421	0.0
Total variance	F10EE	-0.377	-0.289	0.720	-0.524	0.6
Total variance	F11DP	0.535	0.574	0.044	12.120	0.0
Total variance	F12PA	0.38	0.689	0.023	16.857	0.0

Note:

Negative crude residual variance will be highlighted in red

* Crude estimates

† Standardized 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")
)

#display the logics
sem.tab.variance |>
  kable(digits = 3, #rounded
    #format="markdown", #Latex markdown
    booktabs=TRUE, #Latex booktabs
    caption=
      "Residual variance of the dependent variables for initial model",
    escape = F) |> #caption
  kable_styling(latex_options = "striped") |> # gray every other row
  footnote(general =
    "Negative crude residual variance will be highlighted in red",
    symbol = c("Crude estimates",
      "Standardized estimates"))

```


3.2.3 Comment on the result (model 2)

The estimation of Model 2 yielded an overall MLM $\chi^2(426)$ 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 7). 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 ($\text{MLM } \Delta \chi^2[1] = 91.67$), which finalized the decision (Table 7).

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 χ^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 9). 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).

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"))

MI.model2
```

```
##           Parameter      MI    EPC std.EPC
## 1 F12PA → F8SELF 49.859  0.330   0.720
## 2 F5DEC → F12PA 47.088  0.480   0.458
## 3 F8SELF → F10EE 42.770 -1.029  -0.306
##
##                               Logics
## 1      Sensible but not meaningful for this study
## 2                               Sensible and meaningful
## 3 Sensibe and meaningful, but MI is lower than F5→F12
```

Table 10: Select 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

```
#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 = "Select 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 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 indice comparison between the current model (model2) and the initial model was tabulated in table .

```
model3 <- paste(model2, "\nF12PA ~ F5DEC")

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

# Numerical summary of the model:
sem.summary.mlm.a(sem3, 12, 32, "mlm", "Model fit indices for model 2") |>
  unlist()
```

Table 11: Model fit indices for model 2

Measure	Value
chi square	1406.517
df	425.000
p value	0.000
CFI	0.952
TLI	0.944
RMSEA	0.040
RMSEA p value	1.000
SRMR	0.044
CSF	1.117

```
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
sem.compare3 <- rbind(sem.measure1, sem.measure2, sem.measure3) |> data.frame()

sem.compare3.tab <- sem.compare3
sem.compare3.tab$"Δ 2" <- rep(NA, 3)
sem.compare3.tab$"Δ 2"[2] <- chi.diff(sem1, sem2)
```

```
## [1] 91.67
```

```
sem.compare3.tab$"Δ 2"[3] <- chi.diff(sem2, sem3)
```

```
## [1] 46.866
```

```
#turn named vector to data frame
sem.compare3.tab<- sem.compare3.tab %>%
  select(" 2" = chisq.scaled, "DF"=df.scaled,
    "p value"=pvalue.scaled, "Δ 2", "CFI"= cfi.scaled,
    "TLI"= tli.scaled, "RMSEA"= rmsea.scaled,
    "SRMR"=srmr_bentler, "CSF"=chisq.scaling.factor)

sem.compare3.tab[1,4] <- "--"
rownames(sem.compare3.tab) <- c("model1", "model2*", "Model3")

sem.compare3.tab |>
  kable(booktab =T,
    #format = "markdown",
    caption = "Comparison of new and preceding models",
```

Table 12: Comparison of new and preceding models

	2	DF	p value	$\Delta 2$	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

```
align = "r") |>
kable_styling() |>
footnote(symbol = "Model1 + parameter 'F4→F11' set free to estimate")
```

3.3.2 Estimate SEM model 3

3.3.3 Comment on the result (model 3)

3.3.4 Model mis-specification

3.4 Post hos analysis (Model 4)

3.4.1 Compare SEM model 4 with preceding models

3.4.2 Estimate SEM model 4

3.4.3 Comment on the result (model 4)

3.4.4 Model mis-specification

3.5 Post hos analysis (Model 5)

3.5.1 Compare SEM model 5 with preceding models

3.5.2 Estimate SEM model 5

3.5.3 Comment on the result (model 5)

3.5.4 Model mis-specification

3.6 Post hos analysis (Model 6)

3.6.1 Compare SEM model 6 with preceding models

3.6.2 Estimate SEM 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)
```