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

Rong Guang

2023-02-19

Contents

The texts that reflect my understanding have been highlighted in red color.

1 Task description

The first section is task description, which is copied from the assignment5.rmd. It is for communicating with future "me". Please skip it.

1.1 Exercise 5.1

Specify and estimate the initial baseline models for the two groups.

Present a brief summary of the model fit and make the first step of the modification by including **(exceptionally, at the same time!)** all the four parameters known to be required for improving the model fit of both models.

Fine-tune the models step by step following the guidelines given in the lecture material, i.e., implement the modifications **(as usually, one change at a time)** testing and studying each step.

Present the final baseline models of each group and draw the graphs

2 Preparation

##Read in the data set:

Start by downloading the **two data files** from Moodle to your Project folder!

```
#install the necessary pakages
if (!require("pacman")) install.packages("pacman")
pacman::p_load(here,
               expss,
               tidyverse,
               janitor,
               knitr,
               qualtRics,
               arules,
               arulesViz,
               sjlabelled,
               DT,
               stringr,
               labelled,
               ggstatsplot,
               ggcorplot)

library(tidyverse)
library(readr)

#This week's file name
latest.name1 <- "MBIELM1.CSV"
latest.name2 <- "MBISEC1.CSV"
#read in the data
mbi.elm <-  #elementary school
  read_csv(
    file.path(
      here(),
      'data',
      latest.name1
      )
    )

mbi.sec <- #secondary school
  read_csv(
    file.path(
      here(),
      'data',
      latest.name2
      )
    )
```

2.1 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.

2.1.1 To generate a function for calculating chi square difference was defined.

2.1.2 to generate CFA results with improved readability

2.1.3 Write a function to simplify plotting of merged tables for multi-group fit indicies

```
multi.fit.tab <- function(data, title, more.footnote = NULL){
data <- data |>
  rename(p = 'p value',
         p2 = 'RMSEA p value',
         chi = 'chi square') |>
  mutate(df = as.numeric(df) |> round(0),
         p = case_when(
           as.numeric(p) < 0.001 ~ "<0.001",
           as.numeric(p) >= 0.001 ~ p
           ),
         p2 = case_when(
           as.numeric(p2) < 0.001 ~ "<0.001",
           as.numeric(p2) >= 0.001 ~ p2
           )
         ) |>
  mutate('Chi square (df, p)' =
           paste0(chi, "(", df,", ", p, ")"),
         'RMSEA(p)'           =
           paste0(RMSEA, "(", p2, ")"
                  )
         ) |>
  select(
    Model,
    'Chi square (df, p)',
    CFI, TLI,
    'RMSEA(p)',
    SRMR,
    'CSF*'= CSF
    )
#print the combined table with adjustment of aesthetics
data |>
  kable(booktabs = T,
        #format = "markdown",
        caption =
          title,
        align = "lrrrrrr"
        ) |>
  kable_styling(full_width = T) |>
  footnote(symbol =
             c("Chi square scaling factor",
               more.footnote)
           ) |>
  column_spec(1, width = "3.5cm") |>
  column_spec(2, width = "4cm")|>
  column_spec(3, width = "1cm")|>
  column_spec(4, width = "1cm")|>
  column_spec(5, width = "2.5cm")|>
```

```
  column_spec(6, width = "1cm") |>
  column_spec(7, width = "1cm")
}
```

2.1.4 Write a function to simplify plotting aligned residual variance and co-variance tables

```
align.table <- function(data, num.no.header.col, title){

data  |>
  kable(
    digits = 3,
    booktabs = T,
    #format  = "markdown",
    caption = title,
    linesep = ""
    ) |>
  add_header_above(c(" " = num.no.header.col,
                     "Elementary level" = 5,
                     "Secondary level" = 5
                     )
                   ) |>
  kable_styling(
    latex_options = "striped"
  ) |>
  footnote(
           symbol = c(
             "Un-standardized estimates",
             "Standardized estimates"
                      )
           )
}
```

2.1.5 Write a function for correlation matrix with numbers

2.1.6 to generate a function for histogram overlapping with density plot

2.1.7 to generate a function for violin overlapping with box plot

2.1.8 To generate a function describing continuous data set

2.1.9 Write a function describing continuous data set

2.1.10 Write a function for histogram overlapping with density plot

2.1.11 Write a function to generate dot distribution plot

```
dot.dist <-
  function(data, type, title){
    data |>
```

```
      t() |>
      as.data.frame() %>%
      mutate(Item = rownames(.)) |>
      rowwise() |>
      mutate(Median = eval(parse(text = type))(V1:V580)) |>
      ggstatsplot::ggdotplotstats(
        point.args = list(color = "red", size = 3, shape = 13),
        xlab = paste(type, "ratings"),
        title = title,
        x = Median,
        y = Item
      )
    }
```

2.1.12 Write a fuction to generate correlation matrix with statistical test

```
mycor <-
  function(data, cols, title){
  mbi.elm |>
      select(all_of(cols)) |>
      ggstatsplot::ggcorrmat(
        colors = c("#B2182B", "white", "#4D4D4D"),
        title = "(a) Items on emotional exhaustion,
        elementary school teacher",
        matrix.type  = "lower"
      )
    }
```

3 Inspect the data

3.1 Distribution

```
#generate the plots, by subgroup of teachers
p.dist.elm <-
  corr.density(
    mbi.elm,
    fig.num = "1(a)",
    group = "elementary school teacher"
    )
p.dist.sec <-
  corr.density(
    mbi.sec,
    fig.num = "1(b)",
    group = "secondary school teacher"
    )
#print the plot
library(patchwork); p.dist.elm/p.dist.sec
```

Figure 1(a) Distribution of selected items for elementary school teacher

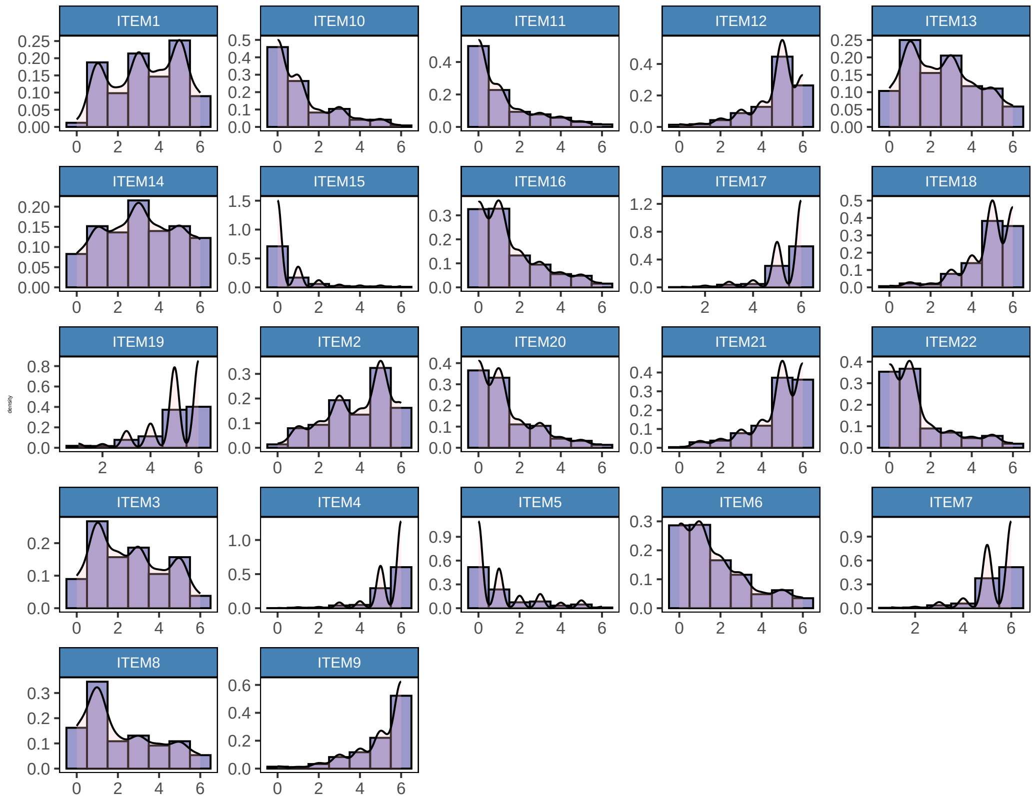
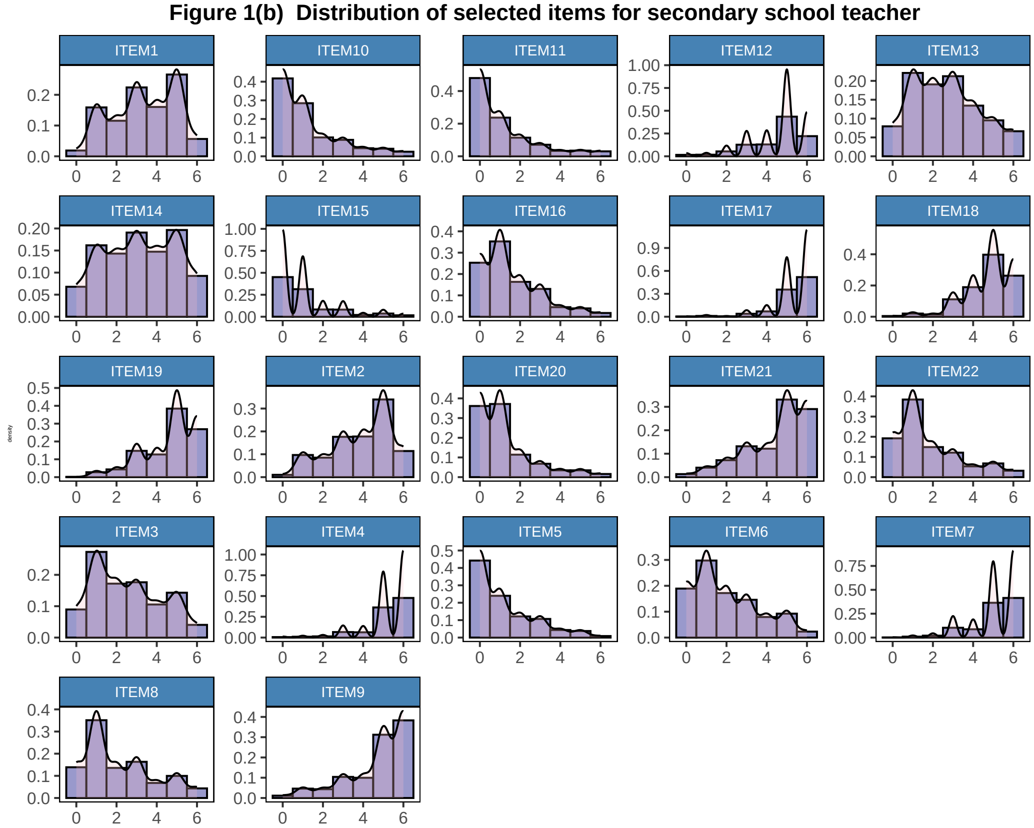
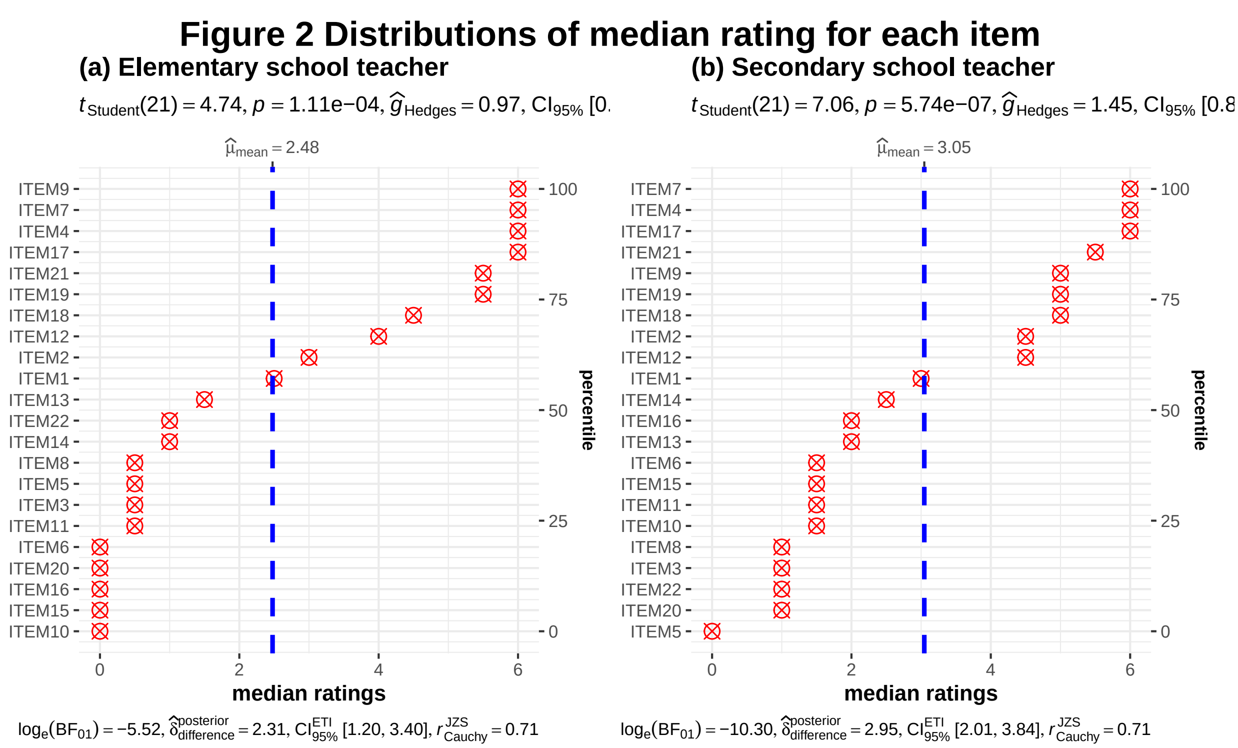


Figure 1(b) Distribution of selected items for secondary school teacher



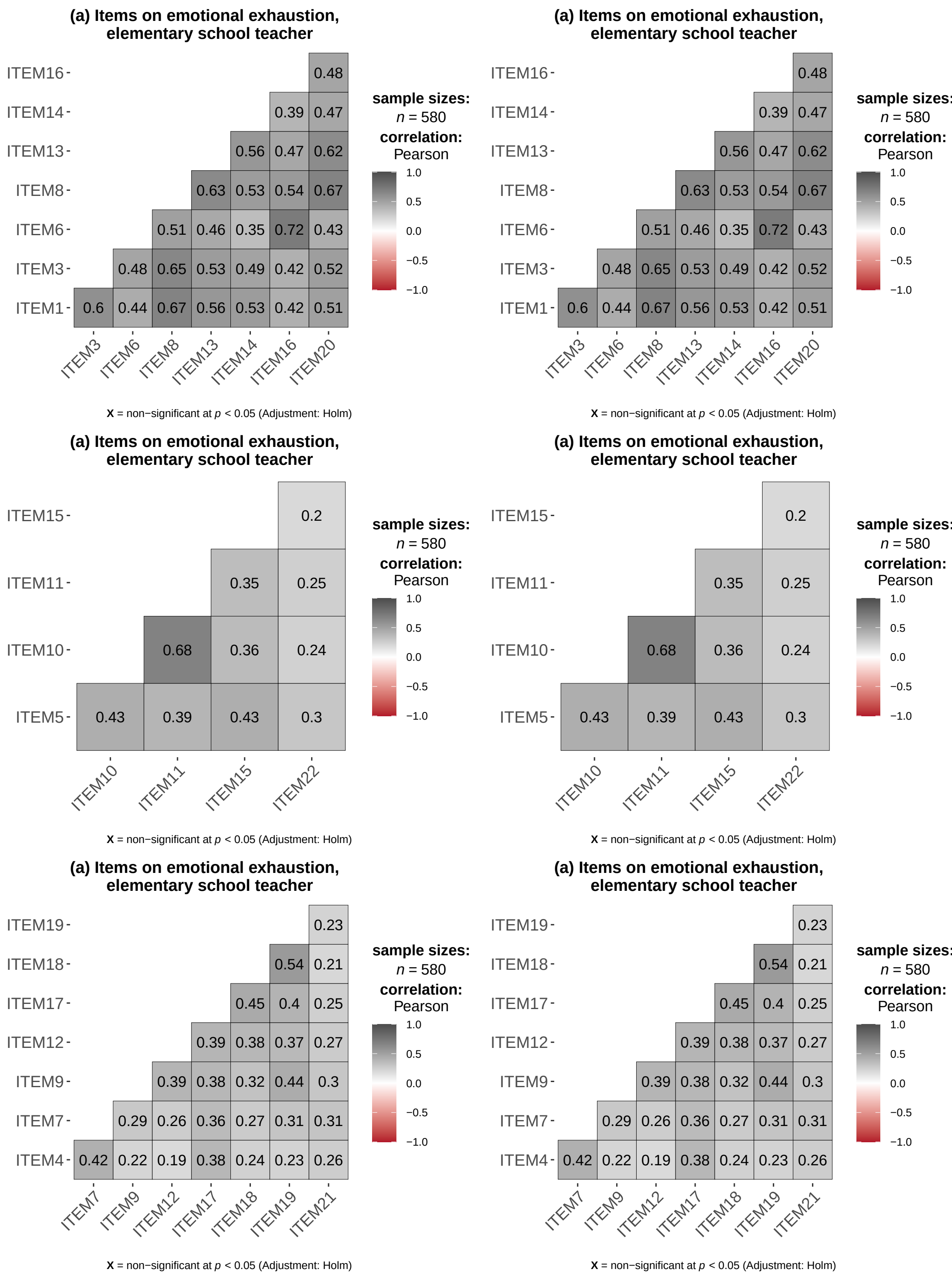
```
#generate plot by subgroups of teachers
p.dot.elm <-
  dot.dist(
    data = mbi.elm,
    type = "median",
    title = "(a) Elementary school teacher"
    )
p.dot.sec <-
  dot.dist(
    data = mbi.sec,
    type = "median",
    title = "(b) Secondary school teacher"
    )
#plot layout
patchwork <- p.dot.elm|p.dot.sec
#print the plot with a genral title
patchwork+plot_annotation(
    title =
      'Figure 2 Distributions of median rating for each item',
    theme =
      theme(plot.title =
              element_text(
                size = 16,
                face = "bold",
                vjust = -1.5,
                hjust =0.5)
            )
    )
```



```
fa.ee <- c("ITEM1", "ITEM3", "ITEM6", "ITEM8", "ITEM13", "ITEM14", "ITEM16", "ITEM20")
fa.dp <- c("ITEM5", "ITEM10", "ITEM11", "ITEM15", "ITEM22")
fa.pa <- c("ITEM4", "ITEM7", "ITEM9", "ITEM12", "ITEM17", "ITEM18", "ITEM19", "ITEM21")
#generate 6 plots, 3 factors X 2 subgroups of teachers
p.cor.elm.ee <-
       mycor(
         data= mbi.elm,
         cols = fa.ee,
         "(a) Items on emotional exhaustion,
         elementary school teacher"
         )
p.cor.sec.ee <-
       mycor(
         data = mbi.sec,
         cols = fa.ee,
         "(b) Items on emotional exhaustion,
          secondary school teacher"
         )
p.cor.elm.dp <-
       mycor(
         data = mbi.elm,
         cols = fa.dp,
         "(c) Items on depersonalization,
          elementary school teacher"
         )
p.cor.sec.dp <-
       mycor(
         data = mbi.sec,
         cols = fa.dp,
         "(d) Items on depersonalization,
          secondary school teacher"
         )
p.cor.elm.pa <-
       mycor(
         data = mbi.elm,
         cols = fa.pa,
         "(e) Items on personal accomplishment,
         secondary school teacher"
         )
p.cor.sec.pa <-
       mycor(
         data = mbi.sec ,
         cols = fa.pa,
         "(f) Items on personal accomplishment,
          secondary school teacher"
         )
#plot sub-figure layout
patchwork <-
  p.cor.elm.ee/p.cor.elm.dp/p.cor.elm.pa|p.cor.sec.ee/p.cor.sec.dp/p.cor.sec.pa
#print the plot with a gernal title
patchwork+
  plot_annotation(
    title =
```

```
    'Figure 3 Correlalogram for items on each factor for two groups of teachers',
  theme =
    theme(plot.title =
            element_text(
              size = 16,
              face = "bold",
              vjust = -1.5,
              hjust =0.5)
          )
  )
```

Figure 3 Correlalogram for items on each factor for two groups of teachers



4 Testing the factorial invariance of MBI inventory between elementary and secondary school teachers

4.1 Define and estimate initial models for both subgroups

The postulated three-factor structure of the MBI that was tested in the previous assignments were re-tested as the initial model for establishing a baseline model.

4.1.1 Define the initial model

```
library(lavaan)
# Define a CFA model using the lavaan (Latent Variable Analysis) syntax:
# see https://lavaan.ugent.be/tutorial/syntax1.html
initial.model <- '
# CFA model for the burnout, the baseline model:
    EE =~ ITEM1 + ITEM2 + ITEM3 + ITEM6 + ITEM8 +
          ITEM13 + ITEM14 + ITEM16 + ITEM20
    DP =~ ITEM5 + ITEM10 + ITEM11 + ITEM15 +ITEM22
    PA =~ ITEM4 + ITEM7 + ITEM9 + ITEM12 +
          ITEM17 + ITEM18 + ITEM19 + ITEM21
          '
```

Cited from Byrne: *It is important to note that measuring instruments are often group specific in the way they operate, and, thus, it is possible that baseline models may not be completely identical across groups.*

4.1.2 Estimate indices to examine factorial validity

(1) Estimate factorial validity for the elementary teacher subgroup

```
cfa.elm <-
  cfa(
    initial.model,
    data = mbi.elm,
    estimator = "MLM",
    mimic = "Mplus"
    )
```

(2) Estimate factorial validity for the secondary teacher subgroup

```
cfa.sec <-
  cfa(
    initial.model,
    data = mbi.sec,
    estimator = "MLM",
    mimic = "Mplus"
    )
```

Table 1: Fit indices for two subgroups, basline models

Model	Chi square (df, p)	CFI	TLI	RMSEA(p)	SRMR	CSF*
Elementary level	826.573(206, <0.001)	0.857	0.840	0.072(<0.001)	0.068	1.225
Secondary level	999.359(206, <0.001)	0.836	0.816	0.075(<0.001)	0.077	1.284

* Chi square scaling factor

4.1.3 Evaluate model

(1) Fit indices

```
library(knitr);library(kableExtra)
#combine fit indices of both levels
initial.elm.fit <-
  cfa.summary.mlm.a(cfa.elm) |>
  t() |>
  as.data.frame()

initial.sec.fit <-
  cfa.summary.mlm.a(cfa.sec) |>
  t() |>
  as.data.frame()

initial.both <-
  rbind(
    initial.elm.fit[2,],
    initial.sec.fit[2,]
    )

names(initial.both) <-
  initial.elm.fit[1,]

rownames(initial.both) <- NULL

initial.both <-
  initial.both |>
  mutate(Model = c("Elementary level",
    "Secondary level")) |>
  select(Model, everything())

#print the table
multi.fit.tab(initial.both, "Fit indices for two subgroups, basline models")
```

See table 1. Goodness-of-fit statistics for this baseline model (three factor) reveals that the indices are less than optimal for both elementary (MLM Chi-square[206] = 826.573; CFI = 0.857; RMSEA = 0.072 ; SRMR = 0.068) and secondary (MLM Chi-square[206] = 999.359; CFI = 0.836; RMSEA = 0.075; SRMR = 0.077) levels.

(2) factor loading

Factor loading of elementary level were extracted.

```
fl.elm <- cfa.summary.b (cfa.elm) #fl  is for factor loading)
colnames(fl.elm)[2] <- "Beta*"
```

Factor loading of secondary level were extracted.

```
fl.sec <- cfa.summary.b (cfa.sec) #fl is for factor loading
colnames(fl.sec) <- c("Parameter",
                     "Beta* ",
                     "SE ",
                     "Z ",
                     "p-value ")
```

Factor loading of both levels were merged in one table and printed.

```
fl.both <- left_join(fl.elm,
                     fl.sec,
                     by = "Parameter")
fl.both |>
  kable(
    digits = 3,
    booktabs = T,
    #format = "markdown",
    caption = "Factor loadings for both levels",
    linesep = ""
    ) |>
  add_header_above(c(" " = 1,
                     "Elementary level" = 4,
                     "Secondary level" = 4
                     )
                   ) |>
  kable_styling() |>
  row_spec(1:9,
           background = "#E5E4E2"
           ) |>
  row_spec(15:22,
           background = "#E5E4E2"
           ) |>
  row_spec(c(1,10,15), bold = T) |>
  footnote(general =
             "Rows with coeffcient estimates fixed to 1 are highligted in bold ",
           symbol = c(
             "Standardized estimates"
                      )
           )
```

the cross-loading involved the loading of Item 12 on Factor 1 (Emotional Exhaustion) in addition to its targeted Factor 3 (Personal Accomplishment)

(3) Variance

Variance of elementary level were extracted.

Table 2: Factor loadings for both levels

	Elementary level				Secondary level			
Parameter	Beta*	SE	Z	p-value	Beta*	SE	Z	p-value
EE→ITEM1	**0.776**	**0.000**	**NA**	**NA**	**0.756**	**0.000**	**NA**	**NA**
EE→ITEM2	0.754	0.032	28.561	<0.001	0.736	0.031	30.236	<0.001
EE→ITEM3	0.740	0.045	21.984	<0.001	0.722	0.043	24.030	<0.001
EE→ITEM6	0.631	0.051	16.064	<0.001	0.626	0.046	18.669	<0.001
EE→ITEM8	0.855	0.042	28.448	<0.001	0.833	0.046	25.968	<0.001
EE→ITEM13	0.754	0.045	22.474	<0.001	0.762	0.045	23.619	<0.001
EE→ITEM14	0.655	0.046	19.939	<0.001	0.634	0.045	20.685	<0.001
EE→ITEM16	0.640	0.047	15.992	<0.001	0.596	0.047	15.261	<0.001
EE→ITEM20	0.734	0.045	18.371	<0.001	0.707	0.048	17.421	<0.001
DP→ITEM5	**0.576**	**0.000**	**NA**	**NA**	**0.453**	**0.000**	**NA**	**NA**
DP→ITEM10	0.794	0.115	11.968	<0.001	0.820	0.188	10.259	<0.001
DP→ITEM11	0.793	0.122	11.588	<0.001	0.808	0.197	9.666	<0.001
DP→ITEM15	0.505	0.072	9.287	<0.001	0.472	0.098	10.295	<0.001
DP→ITEM22	0.351	0.091	6.997	<0.001	0.447	0.131	8.226	<0.001
PA→ITEM4	**0.447**	**0.000**	**NA**	**NA**	**0.340**	**0.000**	**NA**	**NA**
PA→ITEM7	0.516	0.148	7.308	<0.001	0.545	0.221	7.495	<0.001
PA→ITEM9	0.581	0.280	6.629	<0.001	0.681	0.365	7.432	<0.001
PA→ITEM12	0.611	0.303	6.214	<0.001	0.586	0.283	7.398	<0.001
PA→ITEM17	0.681	0.185	7.796	<0.001	0.546	0.187	7.486	<0.001
PA→ITEM18	0.628	0.276	6.628	<0.001	0.698	0.294	7.431	<0.001
PA→ITEM19	0.643	0.255	6.844	<0.001	0.706	0.324	7.565	<0.001
PA→ITEM21	0.425	0.187	7.018	<0.001	0.410	0.242	6.808	<0.001

Note:

Rows with coeffcient estimates fixed to 1 are highligted in bold

* Standardized estimates

```r
var.elm <- cfa.summary.c(cfa.elm, fa.num = 3, item.num = 22)
names(var.elm)[3] <- "Beta*"
names(var.elm)[4]<- "Beta†"
```

Variance of secondary level were extracted.

```r
var.sec <- cfa.summary.c(cfa.sec, fa.num = 3, item.num = 22)
var.sec <- var.sec[,-1]
names(var.sec) <-
  c("Indicator",
    "Beta* ",
    "Beta† ",
    "SE ",
    "Z ",
    "p-value "
    )
```

Variance of both levels were merged in one table and printed.

```r
var.both <- left_join(var.elm,
                      var.sec,
                      by = "Indicator")

align.table(data = var.both,
            num.no.header.col = 2,
            title = "Residual variance for both levels")
```

(3) Co-variance

Co-variance of elementary level were extracted.

```r
cov.elm <- cfa.summary.d(cfa.elm, fa.num = 3, item.num = 22)
colnames(cov.elm)[2:3] <- c("Beta*", "Beta†")
```

Co-variance of secondary level were extracted.

```r
cov.sec <- cfa.summary.d(cfa.sec, fa.num = 3, item.num = 22)
colnames(cov.sec) <- c("Parameter", "Beta* ", "Beta† ", "SE ", "Z ", "p-value ")
```

Co-variance of both levels were merged in one table and printed.

```r
cov.both <- left_join(cov.elm,
                      cov.sec,
                      by = "Parameter")

align.table(data = cov.both,
            num.no.header.col = 1,
            title = "Residual co-variance for both levels")
```

Table 3: Residual variance for both levels

		Elementary level					Secondary level				
Parameter	Indicator	Beta*	Beta†	SE	Z	p-value	Beta*	Beta†	SE	Z	p-value
Residual	ITEM1	1.095	0.398	0.062	17.641	<0.001	1.078	0.429	0.056	19.329	<0.001
Residual	ITEM2	1.067	0.432	0.063	16.832	<0.001	1.071	0.459	0.053	20.373	<0.001
Residual	ITEM3	1.322	0.452	0.089	14.773	<0.001	1.383	0.479	0.083	16.704	<0.001
Residual	ITEM6	1.655	0.602	0.098	16.924	<0.001	1.656	0.609	0.084	19.730	<0.001
Residual	ITEM8	0.886	0.269	0.068	13.044	<0.001	0.890	0.306	0.061	14.560	<0.001
Residual	ITEM13	1.281	0.431	0.087	14.663	<0.001	1.167	0.419	0.075	15.574	<0.001
Residual	ITEM14	1.897	0.571	0.113	16.728	<0.001	1.883	0.599	0.110	17.084	<0.001
Residual	ITEM16	1.363	0.591	0.066	20.746	<0.001	1.353	0.645	0.071	19.024	<0.001
Residual	ITEM20	0.954	0.461	0.093	10.210	<0.001	0.983	0.500	0.057	17.125	<0.001
Residual	ITEM5	1.459	0.669	0.119	12.289	<0.001	1.711	0.795	0.100	17.052	<0.001
Residual	ITEM10	0.806	0.370	0.094	8.530	<0.001	0.803	0.328	0.090	8.944	<0.001
Residual	ITEM11	0.848	0.372	0.101	8.404	<0.001	0.854	0.347	0.095	9.013	<0.001
Residual	ITEM15	0.934	0.745	0.119	7.870	<0.001	1.562	0.778	0.112	13.964	<0.001
Residual	ITEM22	2.086	0.877	0.143	14.538	<0.001	2.052	0.800	0.124	16.598	<0.001
Residual	ITEM4	0.696	0.800	0.066	10.568	<0.001	1.074	0.884	0.104	10.372	<0.001
Residual	ITEM7	0.562	0.734	0.058	9.605	<0.001	0.907	0.703	0.064	14.108	<0.001
Residual	ITEM9	1.176	0.662	0.115	10.247	<0.001	1.194	0.536	0.097	12.297	<0.001
Residual	ITEM12	1.039	0.627	0.079	13.108	<0.001	1.177	0.657	0.076	15.418	<0.001
Residual	ITEM17	0.418	0.536	0.048	8.653	<0.001	0.649	0.701	0.063	10.319	<0.001
Residual	ITEM18	0.894	0.606	0.109	8.170	<0.001	0.703	0.512	0.068	10.329	<0.001
Residual	ITEM19	0.753	0.587	0.062	12.153	<0.001	0.847	0.501	0.080	10.595	<0.001
Residual	ITEM21	1.360	0.819	0.124	10.949	<0.001	1.889	0.832	0.111	17.056	<0.001
Total	EE	1.657	1.000	0.114	14.585	<0.001	1.436	1.000	0.097	14.854	<0.001
Total	DP	0.723	1.000	0.111	6.515	<0.001	0.442	1.000	0.085	5.188	<0.001
Total	PA	0.174	1.000	0.046	3.814	<0.001	0.141	1.000	0.034	4.108	<0.001

* Un-standardized estimates

† Standardized estimates

Table 4: Residual co-variance for both levels

	Elementary level					Secondary level				
Parameter	Beta*	Beta†	SE	Z	p-value	Beta*	Beta†	SE	Z	p-value
EE $\longleftrightarrow$ DP	0.688	0.628	0.075	9.171	<0.001	0.451	0.566	0.057	7.928	<0.001
EE $\longleftrightarrow$ PA	-0.254	-0.473	0.037	-6.952	<0.001	-0.177	-0.393	0.029	-6.193	<0.001

* Un-standardized estimates

† Standardized estimates

4.1.4 Model re-specification

(1) Search for mis-specified parameters

To establish baseline models for both panels of teachers that represent good model fit and parsimony, I further investigated the modification indices of the hypothesized models, respectively for two levels.

MIs of elementary level panel were calculated.

```
#extract needed variables
initial.MI.elm <-
  modindices(cfa.elm,
             standardized = TRUE,
             sort. = TRUE,
             maximum.number = 10)
```

MIs of secondary level panel were calculated.

```
#extract needed variables
initial.MI.sec <-
  modindices(cfa.sec,
             standardized = TRUE,
             sort. = TRUE,
             maximum.number = 10)
```

MI tables with 10 largest MI parameters was printed in descending order of MI. Potential mis-specification of most concerns were highlighted in red.

```
MI.both <- rbind(initial.MI.elm, initial.MI.sec)

MI.both    |>
  mutate(
    op = case_when(op == "~~"~"←→",
                   op == "=~"~"→"),
    Parameter =
           paste(lhs, op, rhs)
         ) |>
  select(Parameter,
         MI = mi,
         EPC = epc,
         "std EPC" = sepc.all
         )|>
  kable(digits = 3,
        booktab = T,
        linesep = "",
        caption =
          "Selected modification indices for determining baseline model") |>
  kable_styling(
    latex_options = "striped"
    ) |>
  row_spec(
    c(1:4, 11:14),
    color = "red"
```

Table 5: Selected modification indices for determining baseline model

	Parameter	MI	EPC	std EPC
Elementary level				
183	ITEM6 $\longleftrightarrow$ ITEM16	180.298	0.893	0.595
120	ITEM1 $\longleftrightarrow$ ITEM2	103.177	0.534	0.494
84	EE $\rightarrow$ ITEM12	81.319	-0.400	-0.400
285	ITEM10 $\longleftrightarrow$ ITEM11	67.743	0.688	0.832
348	ITEM18 $\longleftrightarrow$ ITEM19	43.669	0.279	0.340
323	ITEM4 $\longleftrightarrow$ ITEM7	42.833	0.184	0.294
175	ITEM3 $\longleftrightarrow$ ITEM12	28.187	-0.287	-0.245
275	ITEM5 $\longleftrightarrow$ ITEM15	25.815	0.273	0.234
96	DP $\rightarrow$ ITEM16	25.652	0.459	0.257
185	ITEM6 $\longleftrightarrow$ ITEM5	23.753	0.337	0.217
Secondary level				
1201	ITEM1 $\longleftrightarrow$ ITEM2	171.647	0.627	0.583
2851	ITEM10 $\longleftrightarrow$ ITEM11	135.841	1.181	1.426
1831	ITEM6 $\longleftrightarrow$ ITEM16	127.756	0.686	0.458
841	EE $\rightarrow$ ITEM12	118.156	-0.468	-0.419
2751	ITEM5 $\longleftrightarrow$ ITEM15	77.216	0.580	0.355
296	ITEM11 $\longleftrightarrow$ ITEM15	60.947	-0.485	-0.420
147	ITEM2 $\longleftrightarrow$ ITEM20	53.024	-0.324	-0.316
274	ITEM5 $\longleftrightarrow$ ITEM11	48.297	-0.446	-0.369
339	ITEM9 $\longleftrightarrow$ ITEM19	46.617	0.360	0.358
77	EE $\rightarrow$ ITEM10	45.623	-0.394	-0.302

Note:

Rows highlighted in red are of special concerns

```r
  ) |>
  footnote(general =
             "Rows highlighted in red are of special concerns") |>
  pack_rows(index = c(
    "Elementary level" = 10,
    "Secondary level" = 10
    )
  )
```

See table 5. Three exceptionally large residual co-variances and one cross-loading contributed to the misfit of the model for both teacher panels. The residual co-variances involved Items 1 and 2, Items 6 and 16, and Items 10 and 11; the cross-loading involved the loading of Item 12 on Factor 1 (Emotional Exhaustion) in addition to its targeted Factor 3 (Personal Accomplishment).

In reviewing both the MIs and expected parameter change (EPC) statistics for elementary teachers (table 5, upper part), it is clear that all four parameters are contributing substantially to model misfit, with the residual covariance between Item 6 and Item 16 exhibiting the most profound effect.

We see precisely the same pattern on secondary teachers, albeit the effect would appear to be even more pronounced than it was for elementary teachers. One slight difference between the two groups of teachers regards the impact of these four parameters on model misfit. Whereas the residual covariance between Items 6 and 16 was found to be the most seriously misfitting parameter for elementary teachers; for secondary teachers, the residual covariance between Items 1 and 2 was most pronounced.

(2) Re-specify initial model to model 2

The good practice is relaxing one parameter each time. Nonetheless, according to the knowledge derived from our previous work, I included all four mis-specified parameters in a post-hoc model (common to the groups).

First, the 4 parameters were relaxed in model statement.

```
respecified4 <- 'EE =~ ITEM12
                 ITEM6 ~~ ITEM16
                 ITEM10 ~~ ITEM11
                 ITEM1 ~~ ITEM2
                 '
model2 <- paste(initial.model, respecified4)
```

Then, the model fit were re-estimated for both group, respectively

```
#for elementary
cfa2.elm <-
  cfa(
    model2,
    data = mbi.elm,
    estimator = "MLM",
    mimic = "Mplus"
    )
#for secondary
cfa2.sec <-
  cfa(
    model2,
    data = mbi.sec,
    estimator = "MLM",
    mimic = "Mplus"
    )
```

4.1.5 Examine Model 2

(1) Inspect fit indices of model2 (comparing to initial model)

```
#combine fit indices of both levels
model2.elm.fit <-
  cfa.summary.mlm.a(
    cfa2.elm
    ) |>
  t() |>
  as.data.frame()

model2.sec.fit <-
  cfa.summary.mlm.a(
    cfa2.sec
    ) |>
  t() |>
  as.data.frame()
```

Table 6: Fit indices for two subgroups, model 2, comparing to initial model

Model	Chi square (df, p)	CFI	TLI	RMSEA(p)	SRMR	CSF*
Initial model						
Elementary level	826.573(206, <0.001)	0.857	0.840	0.072(<0.001)	0.068	1.225
Secondary level	999.359(206, <0.001)	0.836	0.816	0.075(<0.001)	0.077	1.284
Model 2						
Elementary level	477.667(202, <0.001)	0.936	0.927	0.049(0.679)	0.050	1.224
Secondary level	587.538(202, <0.001)	0.920	0.909	0.053(0.168)	0.056	1.278

* Chi square scaling factor

```
model2.both <-
  rbind(
    model2.elm.fit[2,],
    model2.sec.fit[2,]
    )

names(model2.both) <- model2.elm.fit[1,]

rownames(model2.both) <- NULL

model2.both <-
  model2.both |>
  mutate(Model = c("Elementary level",
    "Secondary level")) |>
  select(Model, everything())

#combine model 1 and 2 tables
compare12 <- rbind(initial.both, model2.both)

#print the table
multi.fit.tab(compare12,
              "Fit indices for two subgroups, model 2, comparing to initial model") |>
  pack_rows(index = c(
    "Initial model" = 2,
    "Model 2" = 2
  )
  )
```

Estimation of this re-specified model, for each teacher group, yielded greatly improved model fit statistics than initial model. See table 6. However, we should note that several statistics, albeit improved comparing to initial model, still fall below the preferable value. For example, CFI from both groups were <0.95.

(2) Modification indices of model 2

To establish baseline models for both panels of teachers that represent good model fit and parsimony, I further investigated the modification indices of model 2, respectively for two groups, to decide if there was any more model mis-fit and mis-specification

MIs of elementary level panel were calculated.

```r
#extract needed variables
model2.MI.elm <-
  modindices(cfa2.elm,
             standardized = TRUE,
             sort. = TRUE,
             maximum.number = 10)
```

MIs of secondary level panel were calculated.

```r
#extract needed variables
model2.MI.sec <-
  modindices(cfa2.sec,
             standardized = TRUE,
             sort. = TRUE,
             maximum.number = 10)
```

MI tables with 10 largest MI parameters was printed in descending order of MI. Potential mis-specification of most concerns were highlighted in red.

```r
MI2.both <- rbind(model2.MI.elm, model2.MI.sec)

MI2.both    |>
  mutate(
    op = case_when(op == "~~"~"←→",
                   op == "=~"~"→"),
    Parameter =
           paste(lhs, op, rhs)
         ) |>
  select(Parameter,
         MI = mi,
         EPC = epc,
         "std EPC" = sepc.all
         )|>
  kable(digits = 3,
        booktab = T,
        linesep = "",
        caption =
          "Selected modification indices for determining baseline model") |>
  kable_styling(
    latex_options = "striped"
    ) |>
  row_spec(
    c(1:2, 11:12),
    color = "red"
    ) |>
  footnote(general =
             "Rows highlighted in red are of special concerns") |>
  pack_rows(index = c(
    "Elementary level" = 10,
    "Secondary level" = 10
    )
    )
```

Table 7: Selected modification indices for determining baseline model

	Parameter	MI	EPC	std EPC
Elementary level				
323	ITEM4 $\longleftrightarrow$ ITEM7	38.931	0.174	0.284
348	ITEM18 $\longleftrightarrow$ ITEM19	38.744	0.266	0.333
115	PA $\rightarrow$ ITEM14	24.435	0.864	0.205
177	ITEM3 $\longleftrightarrow$ ITEM12	23.978	-0.250	-0.227
227	ITEM13 $\longleftrightarrow$ ITEM12	20.493	0.231	0.211
147	ITEM2 $\longleftrightarrow$ ITEM14	16.441	0.245	0.163
99	DP $\rightarrow$ ITEM16	15.733	0.310	0.197
216	ITEM13 $\longleftrightarrow$ ITEM14	14.838	0.281	0.180
82	EE $\rightarrow$ ITEM11	14.750	0.250	0.206
105	DP $\rightarrow$ ITEM17	12.788	-0.173	-0.188
Secondary level				
821	EE $\rightarrow$ ITEM11	67.177	0.472	0.339
339	ITEM9 $\longleftrightarrow$ ITEM19	43.690	0.355	0.357
276	ITEM5 $\longleftrightarrow$ ITEM15	35.576	0.416	0.310
296	ITEM11 $\longleftrightarrow$ ITEM15	29.016	-0.297	-0.206
247	ITEM16 $\longleftrightarrow$ ITEM20	28.900	0.227	0.201
98	DP $\rightarrow$ ITEM14	22.145	-0.490	-0.239
345	ITEM17 $\longleftrightarrow$ ITEM18	21.583	0.147	0.219
335	ITEM7 $\longleftrightarrow$ ITEM21	21.370	0.247	0.191
346	ITEM17 $\longleftrightarrow$ ITEM19	20.742	-0.159	-0.217
149	ITEM2 $\longleftrightarrow$ ITEM20	20.020	-0.171	-0.162

Note:

Rows highlighted in red are of special concerns

See table 7. In reviewing this information for elementary teachers, we observe two MIs larger than all other MIs (ITEM7 with ITEM4; ITEM19 with ITEM18); both represent residual co-variances. I followed Byrne's step in addressing these parameters. According to Byrne, of the two, only the residual covariance between Items 7 and 4 is substantively viable in that there is a clear overlapping of item content. In contrast, the content of Items 19 and 18 exhibits no such redundancy, and, thus, there is no reasonable justification for including this parameter in a succeeding Model 3.

However, in checking the MI for secondary teachers, the decision was made: more work is needed in establishing an appropriate baseline model. Two parameters were of special concern due to their large MI and substantive meaningfulness. They are Item 11 cross-loads onto factor EE, and item 19 co-varies with item 9. This time I operated by the good practice of specifying one parameter each time. Given the substantially large MI representing the cross-loading of Item 11 on factor EE, this parameter alone was included in our next post-hoc model (Model 3 for secondary teachers).

Byrne noted the reasons for making this decision (to further re-specifying model secondary teachers), which I quoted here for future reflection: (a) The model does not yet reflect a satisfactorily good fit to the data (CFI = 0.920); and (b) in reviewing the MIs in Table 7.2, we observe one very large mis-specified parameter representing the loading of Item 11 on Factor 1 (F1 by ITEM11), as well as another substantially large MI representing a residual covariance between Items 19 and 9, both of which can be substantiated as substantively meaningful parameters.

(3) Model re-specification of model 2 to model 3

```
respecified3 <- 'EE =~ ITEM11
                 ITEM4 ~~ ITEM7
                 '
model3 <- paste(model2, respecified3)
```

Then, the model fit were re-estimated for both group, separately.

```
#for elementary
cfa3.elm <-
  cfa(
    model3,
    data = mbi.elm,
    estimator = "MLM",
    mimic = "Mplus"
    )
#for secondary
cfa3.sec <-
  cfa(
    model3,
    data = mbi.sec,
    estimator = "MLM",
    mimic = "Mplus"
    )
```

4.1.6 Examine Model 3

(1) Inspect fit indices of model3 (comparing to model 2)

```r
#combine fit indices of both levels
model3.elm.fit <-
  cfa.summary.mlm.a(
    cfa3.elm
    ) |>
  t() |>
  as.data.frame()

model3.sec.fit <-
  cfa.summary.mlm.a(
    cfa3.sec
    ) |>
  t() |>
  as.data.frame()

model3.both <-
  rbind(
    model3.elm.fit[2,],
    model3.sec.fit[2,]
    )

names(model3.both) <- model3.elm.fit[1,]

rownames(model3.both) <- NULL

model3.both <-
  model3.both |>
  mutate(Model = c("Elementary level",
    "Secondary level")) |>
  select(Model, everything())

#combine model 1 and 2 tables
compare123 <- rbind(initial.both, model2.both, model3.both)

#print the table
multi.fit.tab(compare123,
              "Fit indices for two subgroups, model 3, comparing to preceding models") |>
  pack_rows(index = c(
    "Initial model" = 2,
    "Model 2" = 2,
    "Model 3" =2
  )
  )
```

See table 8. Results from the estimation of Model 3 for elementary teachers yielded goodness-of-fit statistics that represented a satisfactorily good fit to the data (MLM chi square[201] = 466.722; CFI = 0.939; RMSEA = 0.048; SRMR = 0.050). Although a review of Table 9 (find below) reveals several additional moderately large MIs, for balancing goodness-of-fit and parsimony, the decision was model 3 can serve as the baseline model for elementary teachers.

Results from the estimation of Model 3 for secondary teachers, on the other hand, further substantiated the residual covariance between Items 19 and 9 as representing an acutely mis-specified parameter in the model. Thus, for secondary teachers only, model 4 was put to the test with this residual covariance specified as a freely estimated parameter.

Table 8: Fit indices for two subgroups, model 3, comparing to preceding models

Model	Chi square (df, p)	CFI	TLI	RMSEA(p)	SRMR	CSF*
Initial model						
Elementary level	826.573(206, <0.001)	0.857	0.840	0.072(<0.001)	0.068	1.225
Secondary level	999.359(206, <0.001)	0.836	0.816	0.075(<0.001)	0.077	1.284
Model 2						
Elementary level	477.667(202, <0.001)	0.936	0.927	0.049(0.679)	0.050	1.224
Secondary level	587.538(202, <0.001)	0.920	0.909	0.053(0.168)	0.056	1.278
Model 3						
Elementary level	439.971(200, <0.001)	0.945	0.936	0.045(0.921)	0.049	1.209
Secondary level	530.637(200, <0.001)	0.932	0.921	0.049(0.653)	0.053	1.276

* Chi square scaling factor

(2) Modification indices of model 3

MIs of model 3 for each groups were calculated.

```
#elementary
model3.MI.elm <-
  modindices(cfa3.elm,
             standardized = TRUE,
             sort. = TRUE,
             maximum.number = 10)
#secondary
model3.MI.sec <-
  modindices(cfa3.sec,
             standardized = TRUE,
             sort. = TRUE,
             maximum.number = 10)
```

MI tables with 10 largest MI parameters was printed in descending order of MI. Potential mis-specification of most concerns were highlighted in red.

```
MI3.both <- rbind(model3.MI.elm, model3.MI.sec)

MI3.both    |>
  mutate(
    op = case_when(op == "~~"~"←→",
                   op == "=~"~"→"),
    Parameter =
          paste(lhs, op, rhs)
        ) |>
  select(Parameter,
        MI = mi,
        EPC = epc,
        "std EPC" = sepc.all
        )|>
  kable(digits = 3,
        booktab = T,
        linesep = "",
```

Table 9: Selected modification indices for determining baseline model

	Parameter	MI	EPC	std EPC
Elementary level				
348	ITEM18 $\longleftrightarrow$ ITEM19	32.319	0.246	0.318
116	PA $\rightarrow$ ITEM14	25.330	0.975	0.210
178	ITEM3 $\longleftrightarrow$ ITEM12	23.244	-0.246	-0.224
228	ITEM13 $\longleftrightarrow$ ITEM12	21.113	0.234	0.214
148	ITEM2 $\longleftrightarrow$ ITEM14	16.410	0.244	0.163
100	DP $\rightarrow$ ITEM16	16.023	0.289	0.189
217	ITEM13 $\longleftrightarrow$ ITEM14	14.725	0.280	0.179
106	DP $\rightarrow$ ITEM17	14.422	-0.181	-0.203
326	ITEM4 $\longleftrightarrow$ ITEM17	14.171	0.096	0.174
99	DP $\rightarrow$ ITEM14	11.823	-0.354	-0.193
Secondary level				
339	ITEM9 $\longleftrightarrow$ ITEM19	40.040	0.343	0.350
248	ITEM16 $\longleftrightarrow$ ITEM20	28.249	0.223	0.199
345	ITEM17 $\longleftrightarrow$ ITEM18	22.418	0.150	0.224
346	ITEM17 $\longleftrightarrow$ ITEM19	21.823	-0.164	-0.224
335	ITEM7 $\longleftrightarrow$ ITEM21	20.940	0.243	0.187
85	EE $\rightarrow$ ITEM22	20.315	0.321	0.225
991	DP $\rightarrow$ ITEM14	20.071	-0.404	-0.210
1481	ITEM2 $\longleftrightarrow$ ITEM14	19.909	0.239	0.155
150	ITEM2 $\longleftrightarrow$ ITEM20	18.451	-0.164	-0.155
1161	PA $\rightarrow$ ITEM14	16.677	0.735	0.151

Note:
Rows highlighted in red are of special concerns

```
      caption =
        "Selected modification indices for determining baseline model") |>
  kable_styling(
    latex_options = "striped"
    ) |>
  row_spec(
    c(1:2, 11),
    color = "red"
    ) |>
  footnote(general =
             "Rows highlighted in red are of special concerns") |>
  pack_rows(index = c(
    "Elementary level" = 10,
    "Secondary level" = 10
    )
    )
```

(3) Re-specification of model 3 to model 4 (only for secondary teacher)

The parameter ITEM9 ~~ ITEM19 was relaxed for estimation.

```
respecified4 <- 'ITEM9 ~~ ITEM19
                 '
model4 <- paste(model3, respecified4)
```

Then, the model fit were re-estimated for secondary group, only

```
cfa4.sec <-
  cfa(
    model4,
    data = mbi.sec,
    estimator = "MLM",
    mimic = "Mplus"
    )
```

4.1.7 Examine Model 4

Note that at this point I had taken model 3 as the baseline model for elementary teachers, and model 4 was to achieve the baseline model for secondary teachers.

(1) Inspect fit indices of model4 (comparing to 3)

```
model4.sec.fit <-
  cfa.summary.mlm.a(
    cfa4.sec
    ) |>
  t() |>
  as.data.frame()

names(model4.sec.fit ) <- model4.sec.fit[1,]

model4.sec.fit <- model4.sec.fit [-1,]
model4.sec.fit  <-
  model4.sec.fit  |>
  mutate(Model = "Secondary level") |>
  select(Model, everything())

rownames(model4.sec.fit ) <- NULL

#combine model 1 and 2 tables
model3.both[1,1] <- "Elementary level†"
model4.sec.fit[1,1] <- "Secondary level‡"
compare1234 <-
  rbind(initial.both,
        model2.both,
        model3.both,
        model4.sec.fit )

#print the table
multi.fit.tab(compare1234,
              "Fit indices for two subgroups, model 4, comparing to preceding models",
              c("Baseline model for elementary teachers",
                "Baseline model for secondary teachers")) |>
```

Table 10: Fit indices for two subgroups, model 4, comparing to preceding models

Model	Chi square (df, p)	CFI	TLI	RMSEA(p)	SRMR	CSF*
Initial model						
Elementary level	826.573(206, <0.001)	0.857	0.840	0.072(<0.001)	0.068	1.225
Secondary level	999.359(206, <0.001)	0.836	0.816	0.075(<0.001)	0.077	1.284
Model 2						
Elementary level	477.667(202, <0.001)	0.936	0.927	0.049(0.679)	0.050	1.224
Secondary level	587.538(202, <0.001)	0.920	0.909	0.053(0.168)	0.056	1.278
Model 3						
Elementary level†	439.971(200, <0.001)	0.945	0.936	0.045(0.921)	0.049	1.209
Secondary level	530.637(200, <0.001)	0.932	0.921	0.049(0.653)	0.053	1.276
Model 4						
Secondary level‡	502.530(199, <0.001)	0.937	0.927	0.047(0.863)	0.052	1.273

[*] Chi square scaling factor

[†] Baseline model for elementary teachers

[‡] Baseline model for secondary teachers

```
  pack_rows(index = c(
    "Initial model" = 2,
    "Model 2" = 2,
    "Model 3" =2,
    "Model 4" =1
  )
  ) |>
  row_spec(c(5,7),
           color = "red"
           )
```

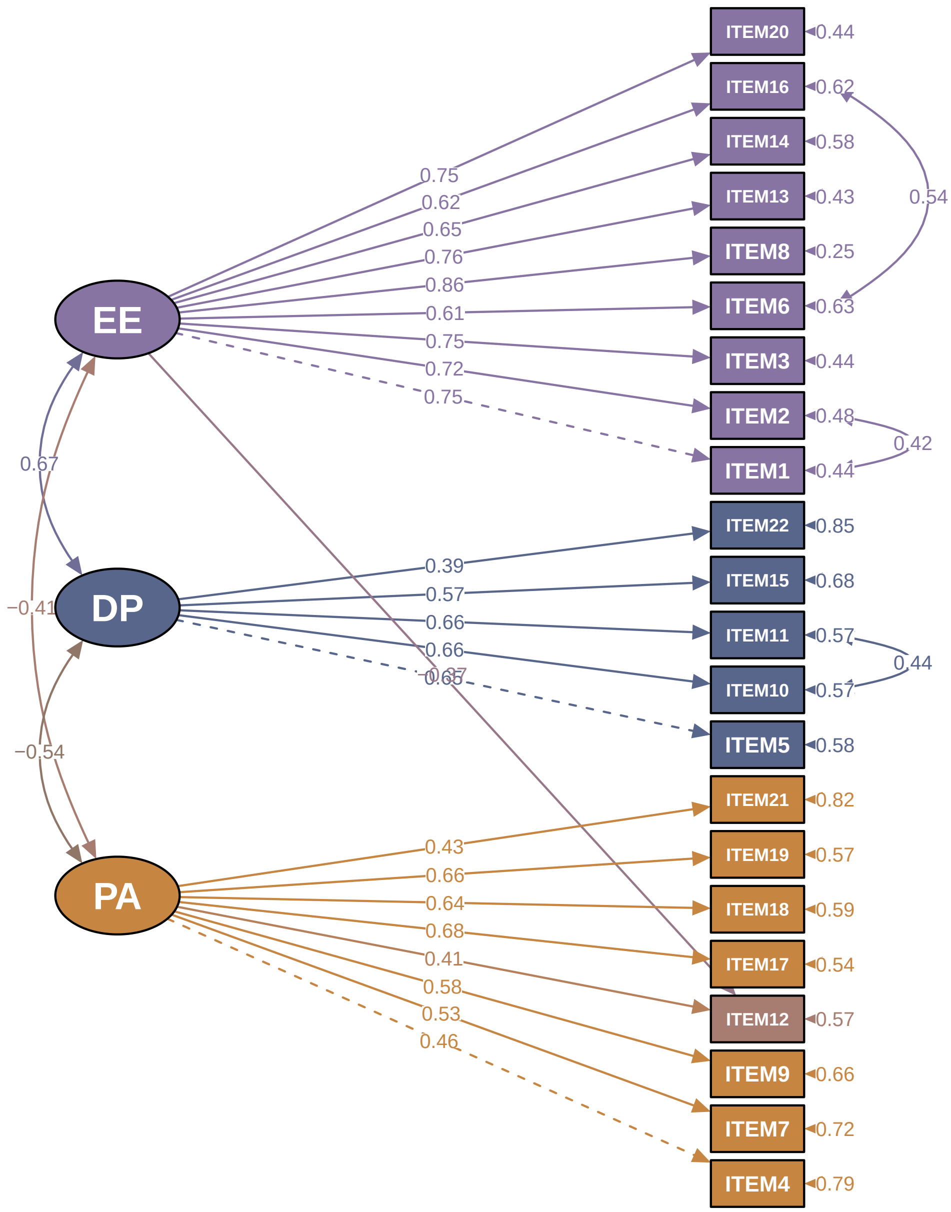
See table 10. Based on a moderately satisfactory goodness-of-fit (MLM cgi-square[200] = 505.831; CFI = 0.937; RMSEA = 0.047; SRMR = 0.052) and to balance fit with parsimony, I consider Model 4 as the final baseline model for secondary teachers.

```
model3 <- 'EE =~ ITEM1 + ITEM2 + ITEM3 + ITEM6 + ITEM8 + ITEM13 + ITEM14 + ITEM16 + ITEM20
           DP =~ ITEM5 + ITEM10 + ITEM11 + ITEM15 +ITEM22
           EE =~ ITEM12
           PA =~ ITEM4 + ITEM7 + ITEM9 + ITEM12 + ITEM17 + ITEM18 + ITEM19 + ITEM21
           ITEM6 ~~ ITEM16
           ITEM10 ~~ ITEM11
           ITEM1 ~~ ITEM2
           '
cfa3 <-
  cfa(
    model3,
    data = mbi.elm,
    estimator = "MLM"
    #mimic = "Mplus"
    )
```

```
library(semPlot)
grps <- list(EE = c("ITEM1", "ITEM2", "ITEM3", "ITEM6", "ITEM8",
        "ITEM13", "ITEM14",  "ITEM16", "ITEM20"),
        DP = c("ITEM5", "ITEM10", "ITEM11", "ITEM15","ITEM22"),
        PA = c("ITEM4", "ITEM7", "ITEM9", "ITEM12",
               "ITEM17", "ITEM18", "ITEM19", "ITEM21"))

order.manifest <- c("ITEM4", "ITEM7", "ITEM9", "ITEM12",
        "ITEM17", "ITEM18", "ITEM19", "ITEM21",
        "ITEM5", "ITEM10", "ITEM11", "ITEM15","ITEM22",
        "ITEM1", "ITEM2", "ITEM3", "ITEM6", "ITEM8",
        "ITEM13", "ITEM14",  "ITEM16", "ITEM20")
order.latent <- c("PA", "DP", "EE")
semPaths(cfa3,
         "col", #un-weighted edges
         "std", #edge label is standarized
         reorder = F,
         latents = order.latent,
         manifest = order.manifest,
         sizeLat = 8,
         sizeLat2 = 5,
         sizeMan = 6,
         sizeMan2 = 3,
         residScale = 4,
         curveAdjacent = "cov", # if edge for adjacent nodes curly or not, "reg"
         shapeMan = "rectangle",
         style = "lisrel",
         group = "latent",

         residuals = T,
         curve = 1.5,
         curvature = 0.2,
         #theme = "colorblind",
         #cardinal = "lat cov",
         curvePivot = F,
         layout = "tree",
         rotation = 2,
         posColor = c("red", "purple", "yellow", "black"),
         color = c("#c68642", "#58668b", "#8874a3"),
         #edge.color = "steelblue",
         shapeLat = "ellipse",
         label.font = 2,
         label.color = "white",
         #Label.scale =T,
         label.prop = 0.7,
         posCol = c("black","black"),
         negCol = c("steelblue","steelblue")
         )
```


EE
DP
PA
ITEM20
ITEM16
ITEM14
ITEM13
ITEM8
ITEM6
ITEM3
ITEM2
ITEM1
ITEM22
ITEM15
ITEM11
ITEM10
ITEM5
ITEM21
ITEM19
ITEM18
ITEM17
ITEM12
ITEM9
ITEM7
ITEM4
0.75
0.62
0.65
0.76
0.86
0.61
0.75
0.72
0.75
0.67
−0.41
−0.54
0.39
0.57
0.66
0.66
0.65
−0.37
0.43
0.66
0.64
0.68
0.41
0.58
0.53
0.46
0.44
0.62
0.58
0.43
0.25
0.63
0.44
0.48
0.44
0.54
0.42
0.85
0.68
0.57
0.57
0.58
0.44
0.82
0.57
0.59
0.54
0.57
0.66
0.72
0.79

```r
chisq_mlm <- function(fit_nested, fit_parent) {
    # scaling correction factors
      c0 <- fitMeasures(fit_nested, "chisq.scaling.factor") %>% as.numeric()
      c1 <- fitMeasures(fit_parent, "chisq.scaling.factor") %>% as.numeric()
    # scaling correction of the difference test
      d0 <- fitMeasures(fit_nested, "df") %>% as.numeric()
      d1 <- fitMeasures(fit_parent, "df") %>% as.numeric()
      cd <- ((d0 * c0) - (d1 * c1))/(d0 - d1)
    # MLM chi-square difference test
      T0 <- fitMeasures(fit_nested, "chisq.scaled") %>% as.numeric()
      T1 <- fitMeasures(fit_parent, "chisq.scaled") %>% as.numeric()
      TRd <- (T0*c0 - T1*c1)/cd
    # degrees of freedom
      df = d0 - d1
    return(c("TR_d" = round(TRd,3),
             "df" = round(df,0),
             "p_value" = pchisq(TRd, df, lower.tail = FALSE) |> round(3)))
}
```

xie