

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

Rong Guang

2023-02-18

Contents

1 Task description	1
1.1 Exercise 5.1	1
2 Preparation	2
2.1 Write functions	2
3 Inspect the data	4
3.1 Distribution	4
4 Testing the factorial invariance of MBI inventory between elementary and secondary school teachers	10
4.1 Define and estimate initial models for both subgroups	10

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 packages
if (!require("pacman")) install.packages("pacman")
pacman::p_load(
  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

###Write a function to generate 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"  
    )  
  )  
}
```

2.1.3 Write a function for correlation matrix with numbers

2.1.4 to generate a function for histogram overlapping with density plot

2.1.5 to generate a function for violin overlapping with box plot

2.1.6 To generate a function describing continuous data set

2.1.7 Write a function describing continuous data set

2.1.8 Write a function for histogram overlapping with density plot

2.1.9 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
)
}

```

2.1.10 Write a function 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

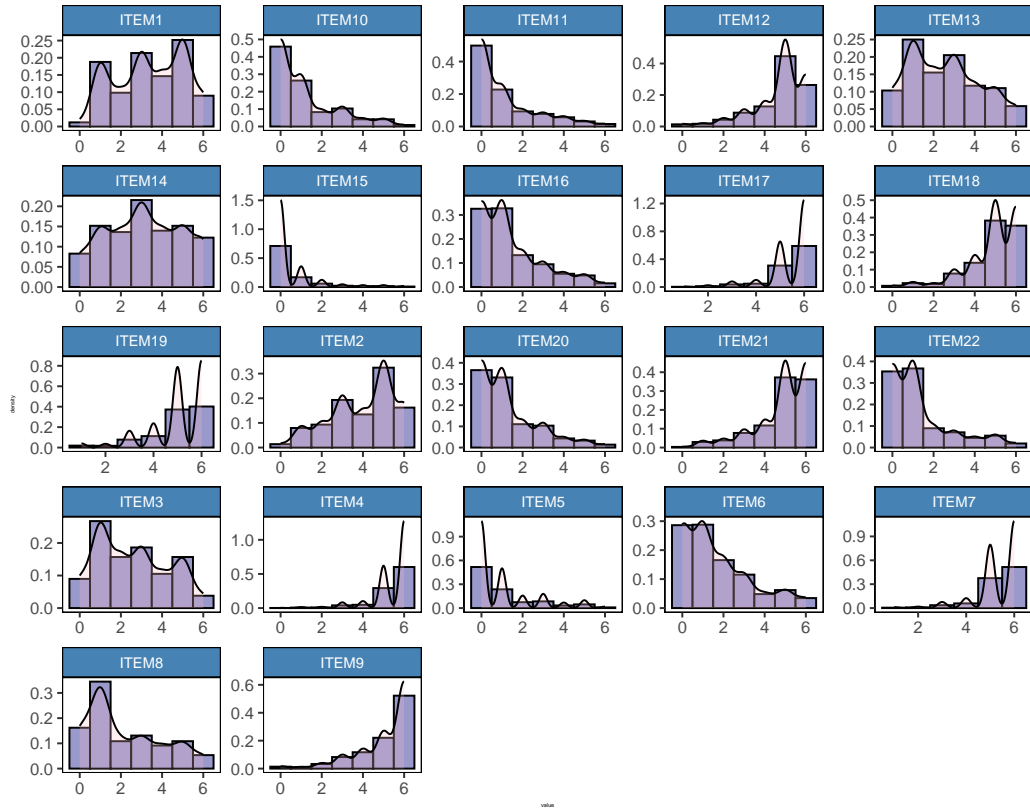
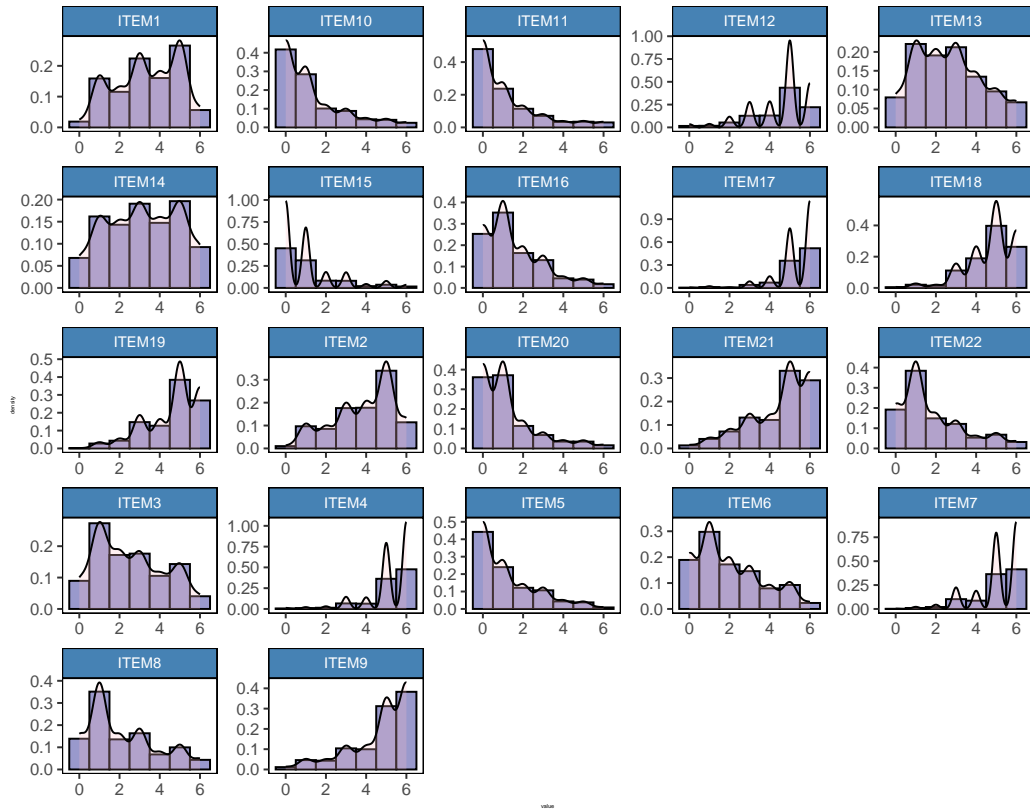


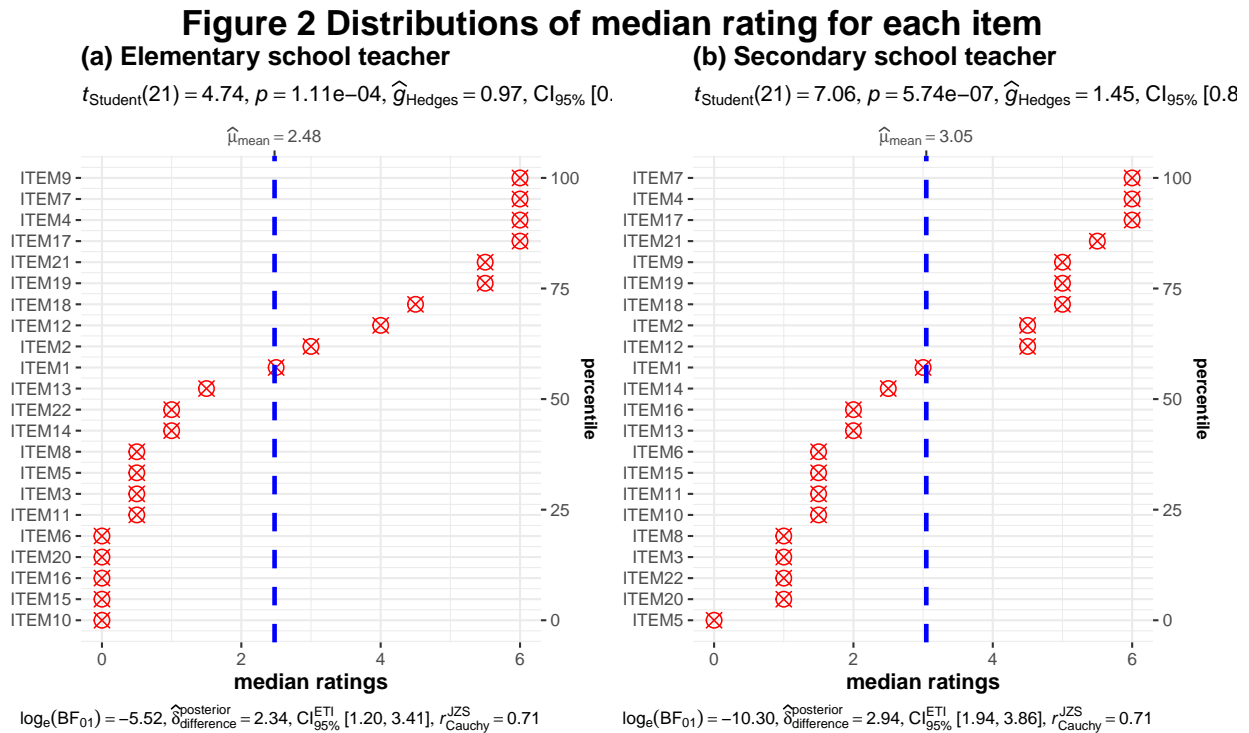
Figure 1(b) Distribution of selected items for secondary school teacher



```

#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 =

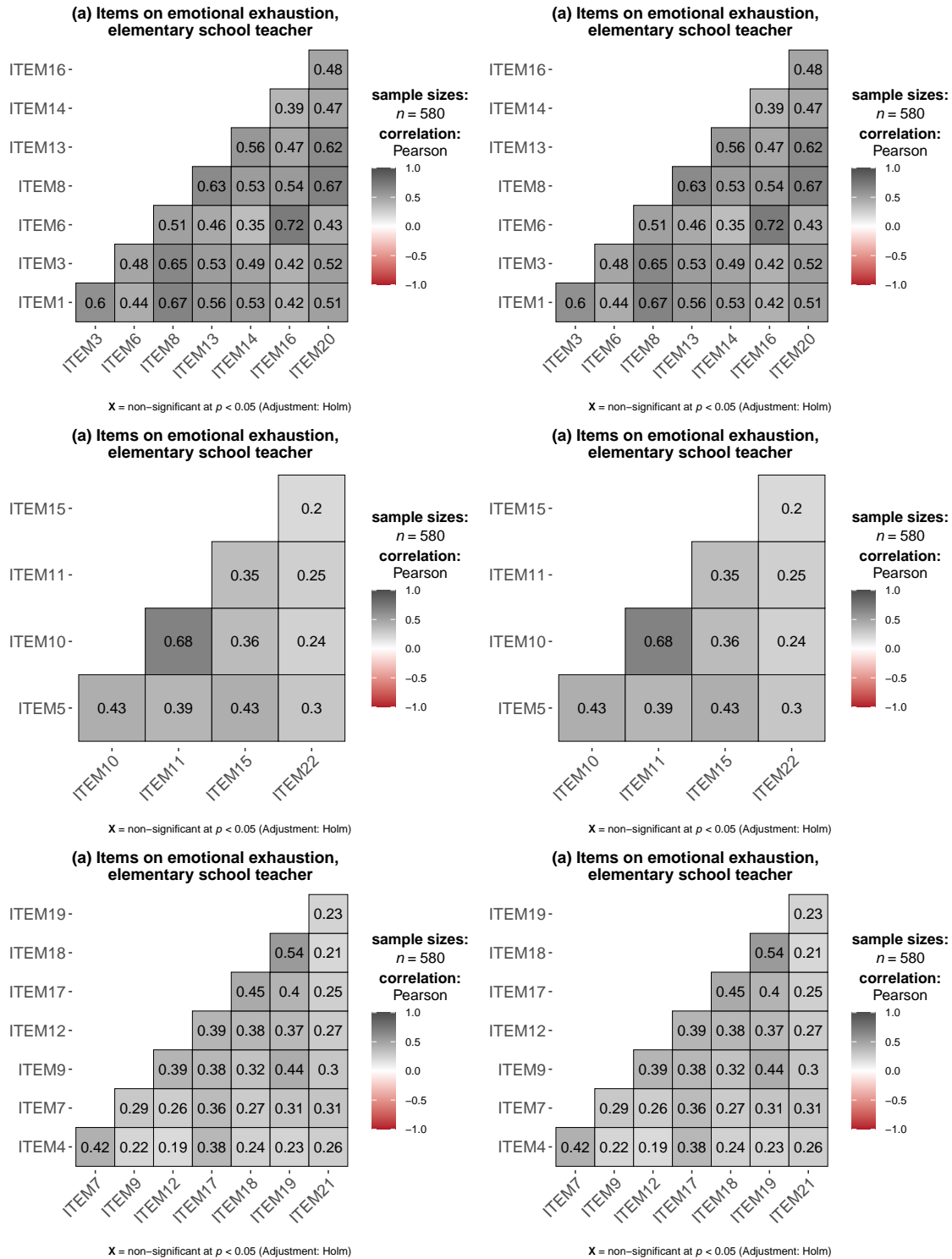
```

```

    '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 model under test in this initial multi-group application is the same postulated three-factor structure of the MBI that was tested in the previous assignments.

4.1.1 Define the baseline model

```
library(lavaan)
# Define a CFA model using the lavaan (Latent Variable Analysis) syntax:
# see https://lavaan.ugent.be/tutorial/syntax1.html
BLmodel <- '
# 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
'
```

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 factorial validity

- (1) Estimate factorial validity for the elementary teacher subgroup

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

- (2) Estimate factorial validity for the secondary teacher subgroup

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

4.1.3 Evaluate model

(1) Fit indices

```
library(knitr);library(kableExtra)
#combine fit indices of both levels
BL.elm.fit <- cfa.summary.mlm.a(cfa.elm) |> t() |> as.data.frame()
BL.sec.fit <- cfa.summary.mlm.a(cfa.sec) |> t() |> as.data.frame()
BL.both <- rbind(BL.elm.fit[2,], BL.sec.fit[2,])
names(BL.both) <- BL.elm.fit[1,]
rownames(BL.both) <- c("Elementary level",
                      "Secondary level")
#improve the readability of the combined table
BL.both <- BL.both |>
  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('Chi square (df, p)',
        CFI, TLI,
        'RMSEA(p)',
        SRMR,
        'CSF*' = CSF
  )
#print the combined table with adjustment of aesthetics
BL.both |>
  kable(booktabs = T,
        #format = "markdown",
        caption =
          "Fit indices for two subgroups, baseline models",
        align = "r"
  ) |>
  kable_styling(full_width = T) |>
  footnote(symbol =
    "Chi square scaling factor"
  ) |>
  column_spec(1, width = "3cm") |>
  column_spec(2, width = "4cm") |>
  column_spec(3, width = "1cm") |>
```

Table 1: Fit indices for two subgroups, baseline models

	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

```
column_spec(4, width = "1cm")|>
column_spec(5, width = "2.5cm")|>
column_spec(6, width = "1cm") |>
column_spec(7, width = "1cm")
```

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
))
```

Table 2: Factor loadings for both levels

Parameter	Elementary level				Secondary level			
	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 coefficient estimates fixed to 1 are highlighted in bold

* Standardized estimates

```

) |>
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 coefficient estimates fixed to 1 are highlighted 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.

```
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.

```
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.

```
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.

```
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.

```
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.

```
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

Parameter	Indicator	Elementary level					Secondary level				
		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

Parameter	Elementary level					Secondary level				
	Beta*	Beta†	SE	Z	p-value	Beta*	Beta†	SE	Z	p-value
EE $\leftarrow \rightarrow$ DP	0.688	0.628	0.075	9.171	<0.001	0.451	0.566	0.057	7.928	<0.001
EE $\leftarrow \rightarrow$ 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

MIIs of elementary level panel were calculated.

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

MIIs of secondary level panel were calculated.

```
#extract needed variables
BL.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(BL.MI.elm, BL.MI.sec)

MI.both |>
  mutate(Parameter =
    paste(rhs, "→", lhs)
  ) |>
  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"
  ) |>
  footnote(general =
    "Rows highlighted in red are of special concerns") |>
  pack_rows(index = c(
    "Elementary level" = 10,
    "Secondary level" = 10
  )
  )
```


Table 5: Selected modification indices for determining baseline model

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

Note:

Rows highlighted in red are of special concerns

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)

xie