COS-D419 Factor Analysis and Structural Equation Models 2023, Assignment 2

Rong Guang

2023-01-31

1 Exercise 2.1

Specify and test the hypothesis given on the page 1 of the lecture material.

Draw conclusions based on the χ^2 statistic and the CFI, TLI, RMSEA, and SRMR indices.

What can you say about the parameter estimates?

Visualize the model.

1.1 Read in the data set

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

```
library(tidyverse)#data wrangling
library(readr)# read data into r
orig_data <- read_csv("ASC7INDM.CSV", show_col_types = FALSE)
```

1.2 Write functions

```
unique.levels <-  function(sc){
  values <- lapply(sc, function(x)sort(unique(x)))
for(x in 1:ncol(sc)){
  a <- paste(c("Variable ",
               names(values)[x],
               " has values of ",
               paste(values[[x]],
                     collapse = ",")),
             collapse = "")
  print(a)
  }
}
```

1.3 Subset the data set

Subset the variables for analysis and name it as sc (Self-concept).

```r
# Select the variables for use
sc <- orig_data %>% dplyr::select(starts_with("SDQ2N")) # namming logic: sc = self-concept
```

1.4 Inspect the data

Have a quick overview of the data.

```r
glimpse(sc)
```

```
## Rows: 265
## Columns: 16
## $ SDQ2N01 <dbl> 6, 6, 4, 5, 6, 5, 1, 2, 5, 4, 2, 5, 6, 4, 4, 6, 6, 6, 5, 6, 6,~
## $ SDQ2N13 <dbl> 5, 6, 6, 5, 5, 5, 6, 1, 5, 6, 6, 5, 6, 3, 5, 6, 6, 6, 4, 5, 5,~
## $ SDQ2N25 <dbl> 4, 6, 6, 5, 5, 5, 1, 6, 6, 3, 6, 6, 6, 5, 5, 6, 6, 6, 6, 5, 4,~
## $ SDQ2N37 <dbl> 6, 6, 2, 6, 4, 3, 6, 4, 6, 6, 6, 5, 5, 5, 4, 5, 6, 4, 4, 6, 6,~
## $ SDQ2N04 <dbl> 3, 6, 6, 5, 3, 3, 4, 4, 6, 6, 5, 6, 5, 4, 4, 4, 4, 6, 5, 5, 3,~
## $ SDQ2N16 <dbl> 4, 6, 4, 6, 4, 2, 6, 4, 6, 5, 6, 6, 5, 5, 5, 5, 6, 5, 4, 6, 6,~
## $ SDQ2N28 <dbl> 4, 6, 6, 5, 4, 4, 6, 4, 6, 6, 6, 6, 5, 5, 5, 5, 6, 4, 2, 4, 4,~
## $ SDQ2N40 <dbl> 6, 6, 3, 6, 4, 4, 6, 6, 6, 6, 6, 6, 6, 5, 4, 4, 6, 6, 5, 5, 5,~
## $ SDQ2N10 <dbl> 2, 5, 6, 5, 4, 4, 1, 6, 5, 4, 2, 6, 5, 5, 5, 3, 4, 6, 5, 4, 6,~
## $ SDQ2N22 <dbl> 6, 6, 5, 6, 6, 4, 6, 6, 6, 6, 6, 6, 6, 5, 6, 6, 6, 6, 6, 3, 6,~
## $ SDQ2N34 <dbl> 1, 6, 4, 3, 5, 5, 1, 1, 5, 4, 5, 6, 5, 2, 5, 2, 3, 2, 1, 3, 3,~
## $ SDQ2N46 <dbl> 5, 6, 5, 5, 6, 6, 6, 5, 6, 6, 6, 6, 6, 6, 2, 5, 6, 6, 6, 6, 6,~
## $ SDQ2N07 <dbl> 6, 6, 6, 6, 3, 4, 5, 3, 6, 5, 6, 6, 6, 6, 4, 4, 6, 6, 6, 6, 3,~
## $ SDQ2N19 <dbl> 6, 6, 6, 6, 4, 5, 6, 4, 6, 6, 5, 6, 6, 6, 5, 5, 6, 6, 5, 5, 5,~
## $ SDQ2N31 <dbl> 6, 6, 3, 6, 4, 4, 6, 4, 6, 6, 6, 6, 6, 6, 5, 5, 6, 6, 5, 5, 5,~
## $ SDQ2N43 <dbl> 6, 6, 1, 5, 5, 4, 5, 6, 6, 6, 6, 6, 6, 6, 5, 6, 6, 6, 5, 6, 5,~
```

The data set includes 16 variables from 265 observations. All the variables are numeric. Next, I examined the unique values of each variables.

```r
unique.levels(sc)
```

```
## [1] "Variable SDQ2N01 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N13 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N25 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N37 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N04 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N16 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N28 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N40 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N10 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N22 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N34 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N46 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N07 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N19 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N31 has values of 1,2,3,4,5,6"
## [1] "Variable SDQ2N43 has values of 1,2,3,4,5,6"
```

For each variable, the values distribute from 1 to 6.

2 Explore the data

2.1 Descriptive statistics

```
library(kableExtra)#improved table visuals
library(psych)#for function "describe"
sc.ds <- sc %>%  #sc.ds = self-concept descriptive statistics
  describe(IQR = T) %>%
    as.data.frame() %>%
  dplyr::select(mean, median, sd, range, se, IQR)
#print the descriptive statistics table
sc.ds %>%
  kable(booktabs=T,
        longtable=T,
        digits = 2,
        caption = "Descriptive dtatistics of selected variables",
        linesep = "") %>%
  add_header_above(c("", "centralized tendency" = 2, "dispersion tendency" = 4)) %>%
  kable_styling(latex_options = c("striped","repeat_header")) %>%
  column_spec(1, width = "3cm", bold = T)
```

Table 1: Descriptive dtatistics of selected variables

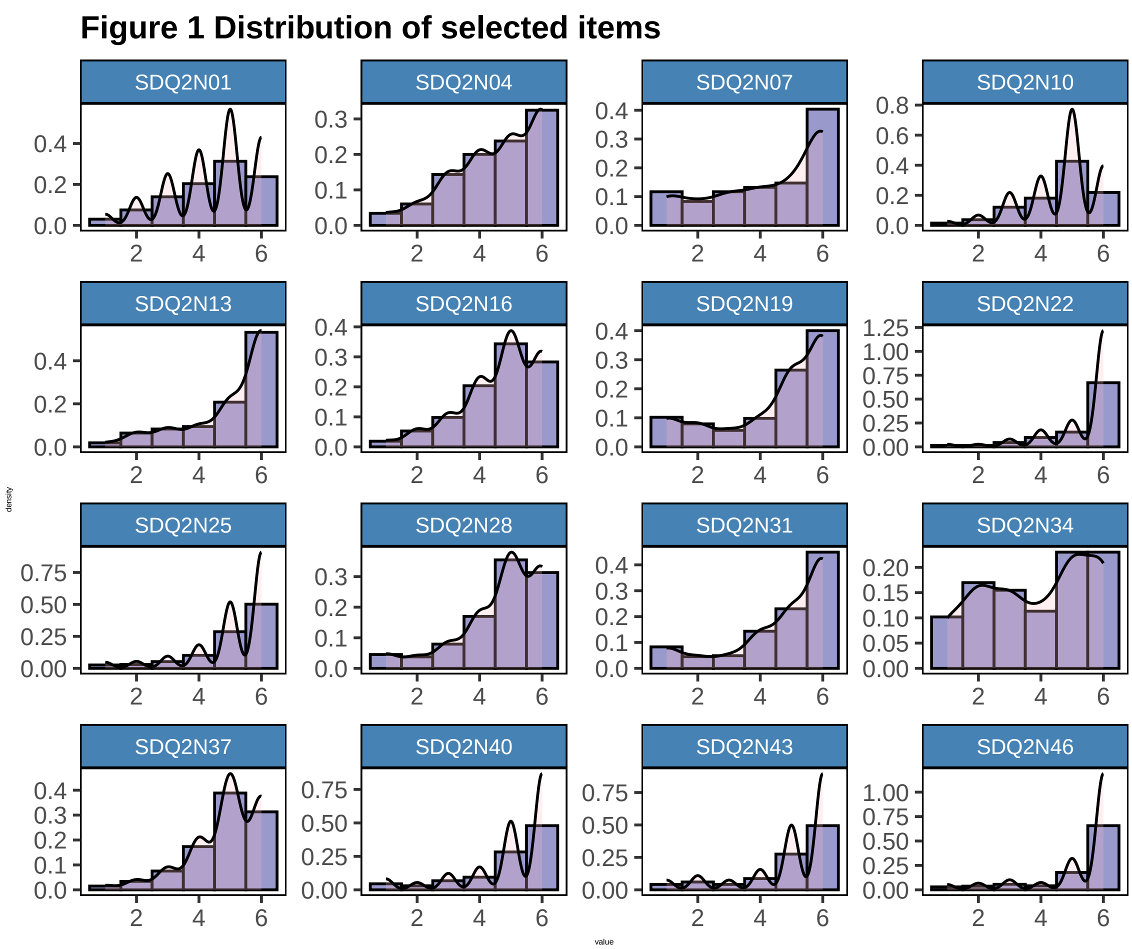
	centralized tendency		dispersion tendency			
	mean	median	sd	range	se	IQR
SDQ2N01	4.41	5	1.35	5	0.08	1
SDQ2N13	5.00	6	1.36	5	0.08	2
SDQ2N25	5.10	6	1.23	5	0.08	1
SDQ2N37	4.83	5	1.14	5	0.07	2
SDQ2N04	4.52	5	1.40	5	0.09	2
SDQ2N16	4.65	5	1.24	5	0.08	2
SDQ2N28	4.69	5	1.33	5	0.08	2
SDQ2N40	4.98	5	1.36	5	0.08	1
SDQ2N10	4.62	5	1.15	5	0.07	1
SDQ2N22	5.38	6	1.09	5	0.07	1
SDQ2N34	3.89	4	1.70	5	0.10	3
SDQ2N46	5.27	6	1.30	5	0.08	1
SDQ2N07	4.32	5	1.78	5	0.11	3
SDQ2N19	4.54	5	1.69	5	0.10	2
SDQ2N31	4.74	5	1.57	5	0.10	2
SDQ2N43	4.98	5	1.40	5	0.09	1

2.2 Visualization

2.2.1 Distribution of each item

The distribution was first examined by histogram.

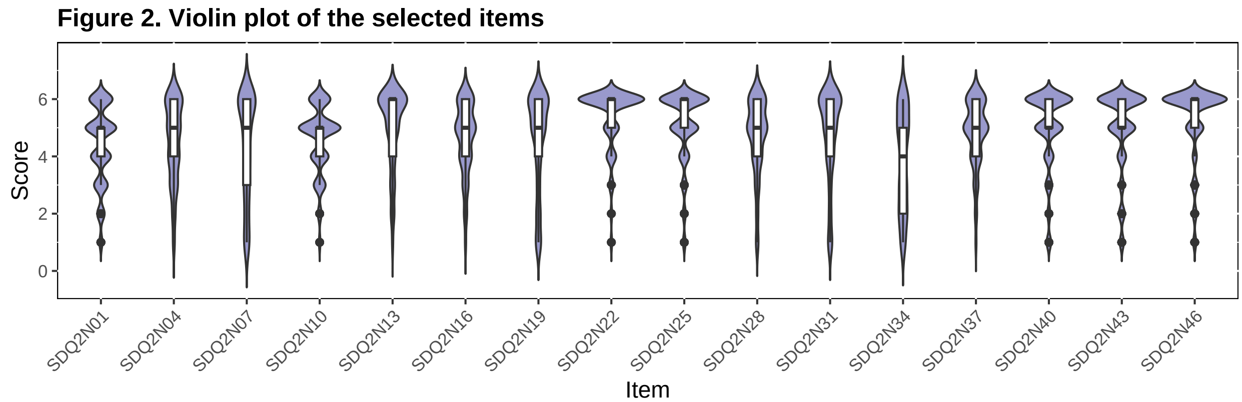
```r
sc %>%
  pivot_longer(everything()) %>%  #longer format
  ggplot(aes(x = value)) + #x axis used variable "value" (a default of pivot)
  geom_histogram(binwidth = 1, aes(y = ..density..), #match ys of density and histogram plots
                 color = "black",  fill = "#9999CC")+  # adjust aesthetics for hist
  geom_density(fill = "pink", alpha = 0.25)+ #adjust aesthetics for density plot
  facet_wrap(~name, scales = "free") + #wrap by name variable
  theme(panel.grid.major = element_blank(), #get rid of the  grids
        panel.grid.minor = element_blank(),
        panel.background = element_rect(fill = "white",#adjust the background
                                        color = "black"),
        strip.background = element_rect(color = "black",#adjust the strips aes
                                        fill = "steelblue"),
        strip.text = element_text(size =8, color = "white"), #adjust strip text
        axis.title.x = element_text(size = 3), #adjust the x text
        axis.title.y = element_text(size = 3), # adjust the y text
        plot.title = element_text(size = 12, face = "bold"))+ #adjust the title
  labs(title = "Figure 1 Distribution of selected items") #title it
```



Most item values skewed to the right, except for SDQ2N34, the values of which were more evenly assigned.

```
sc.long <- sc %>% pivot_longer(everything(), names_to = "item", values_to = "score")

sc.long %>%
  ggplot(aes(x = item, y = score)) +
  geom_violin(trim=F, fill = "#9999CC") +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 45, hjust =1),
        axis.title = element_text(size = 12),
        panel.background = element_rect(fill = "white", color = "black"),
        plot.title = element_text(face="bold"))+
  labs(x = "Item",
       y = "Score",
       title = "Figure 2. Violin plot of the selected items")+
  geom_boxplot(width = 0.1, fill = "white")
```

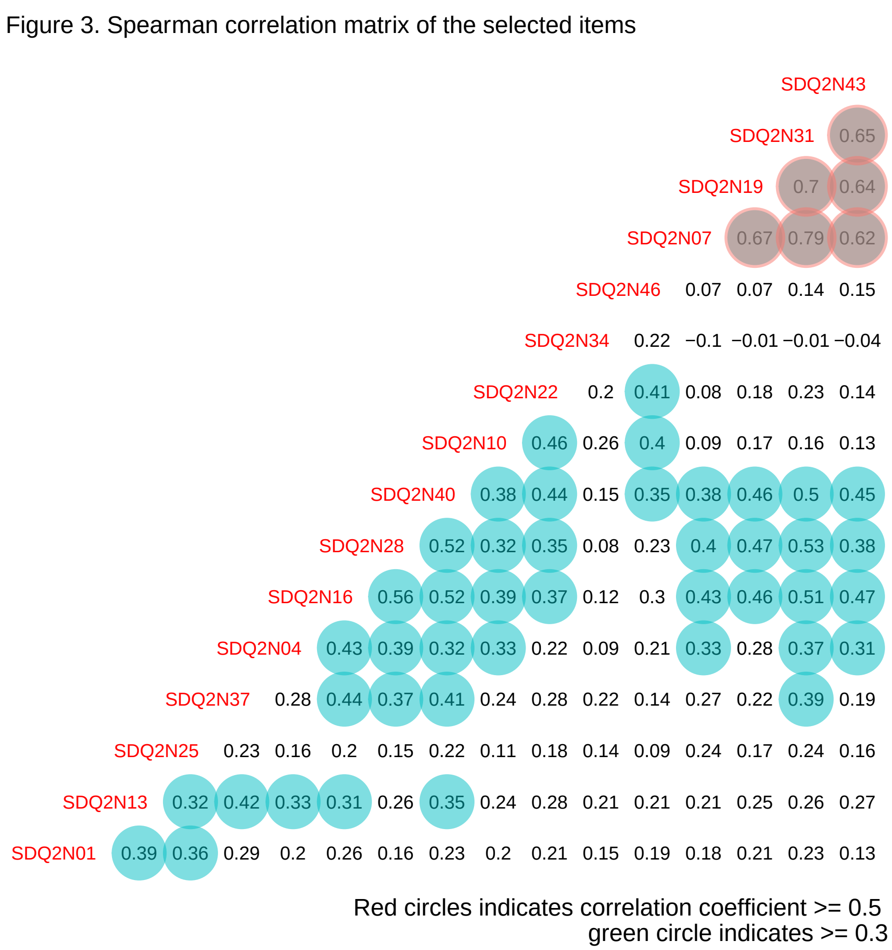


Distribution displayed by violin plot was consistent with histogram, with a couple of outliers observed in SDQ2N01, SDQ2N10, SDQ2N22, SDQ2N25, SDQ2N40, SDQ2N43, SDQ2N46.

2.2.2 Correlation among items

```
library(GGally)
ggcorr(sc,
       geom = "blank",
       label = TRUE,
       hjust = 0.9,
       color = "red",
       face = "bold",
       method = c("pairwise","spearman"),
       digits = 2,
       size= 2.5,
       label_size = 2.5,
       label_round = 2,
       layout.exp =1) +
  geom_point(size = 9,
             aes(color = "red",
                 alpha = abs(coefficient) > 0.3)) +
  scale_alpha_manual(values = c("TRUE" = 0.3, "FALSE" = 0)) +
    geom_point(size = 10,
```

```
                aes(color = "green", alpha = abs(coefficient) > 0.6)) +
  scale_alpha_manual(values = c("TRUE" = 0.5, "FALSE" = 0)) +
  guides(color = FALSE,
         alpha = FALSE) +
  labs(title = "Figure 3. Spearman correlation matrix of the selected items",
       caption =
        " Red circles indicates correlation coefficient >= 0.5
       green circle indicates >= 0.3")+
  theme(plot.title = element_text(size = 9))
```



It is found that each variable correlated with at least one of the other variable with a Spearman correlation coefficient $>= 0.3$, except for items SDQ2N46, SDQ2N25 and SDQ2N34.

2.3 Tests for normality

2.3.1 univariate normality

```
library(MVN)
#univariate test
result <- mvn(sc, mvnTest = "mardia", univariateTest = "SW")
```

Table 2: Univariate distribution tests

Test	Variable	Statistic	p value	Normality
Shapiro-Wilk	SDQ2N01	0.8911	<0.001	NO
Shapiro-Wilk	SDQ2N13	0.7454	<0.001	NO
Shapiro-Wilk	SDQ2N25	0.7357	<0.001	NO
Shapiro-Wilk	SDQ2N37	0.8380	<0.001	NO
Shapiro-Wilk	SDQ2N04	0.8729	<0.001	NO
Shapiro-Wilk	SDQ2N16	0.8666	<0.001	NO
Shapiro-Wilk	SDQ2N28	0.8292	<0.001	NO
Shapiro-Wilk	SDQ2N40	0.7451	<0.001	NO
Shapiro-Wilk	SDQ2N10	0.8657	<0.001	NO
Shapiro-Wilk	SDQ2N22	0.6333	<0.001	NO
Shapiro-Wilk	SDQ2N34	0.8883	<0.001	NO
Shapiro-Wilk	SDQ2N46	0.6232	<0.001	NO
Shapiro-Wilk	SDQ2N07	0.8254	<0.001	NO
Shapiro-Wilk	SDQ2N19	0.7883	<0.001	NO
Shapiro-Wilk	SDQ2N31	0.7705	<0.001	NO
Shapiro-Wilk	SDQ2N43	0.7313	<0.001	NO

Table 3: Multivariate distribution tests

Test	Statistic	p value	Result
Mardia Skewness	2676.70326551468	2.40748109101252e-196	NO
Mardia Kurtosis	35.5380362660965	0	NO
MVN	NA	NA	NO

```
#tabulate the result
result$univariateNormality %>%
  kable(booktabs = TRUE,
        digits = 2,
        caption = "Univariate distribution tests") %>%
  kable_styling(latex_options = "striped") %>%
  row_spec(0, background = "#9999CC")
```

The test showed none of the variables had a normal distribution, which was consistent with the subjective impression.

```
result$multivariateNormality %>%
  kable(booktabs = TRUE,
        digits = 2,
        caption = "Multivariate distribution tests",
        linesep = "") %>%
  kable_styling(latex_options = "striped") %>%
  row_spec(0,background = "#9999CC")
```

The test showed the selected manifest variables did not have multivariate normality.

3 Hypothesis testing

3.1 Self-concept (SC) is a multi-dimensional construct composed of four factors (GSC, ASC, ESC and MSC)

3.1.1 Model estimation

Four factor CFA was performed herein.

```
library(lavaan) #SEM

#define model
model1 <- '# CFA model of self-concept (SC):
           GSC =~ SDQ2N01 + SDQ2N13 + SDQ2N25 + SDQ2N37
           ASC =~ SDQ2N04 + SDQ2N16 + SDQ2N28 + SDQ2N40
           ESC =~ SDQ2N10 + SDQ2N22 + SDQ2N34 + SDQ2N46
           MSC =~ SDQ2N07 + SDQ2N19 + SDQ2N31 + SDQ2N43
          '

# Estimate the model using the data defined earlier
cfa1 <- cfa(model1, data = sc)
```

Measures of goodness of fit and subjective indices of fit were obtained.

```
#turn off scientific notion
options(scipen = 999)

#obtain CFA measures
cfa1.measure <- fitMeasures(cfa1,  #obtain fit measures
                            c("chisq",  #specify selected measures
                              "df",
                              "pvalue",
                              "cfi",
                              "tli",
                              "rmsea",
                              "rmsea.pvalue",
                              "srmr"))

#turn named vector to data frame
cfa1.fig.a <- cfa1.measure %>%
  tibble(name= names(cfa1.measure), value = cfa1.measure) %>% #vector to df
  select(Measure = name, Value = value) %>%  #select and rename columns
  mutate(Value = round(as.numeric(Value),3)) %>%  #round
  kable(format = "markdown", #table aesthetics
        booktabs = T, #Latex table with booktabs
        caption =  #caption
          "Goodness-of-fit and subjective indices of fit for 4 factor CFA") %>%
  kable_styling(latex_options = "striped") %>%
  row_spec(0, background = "#9999CC") #color first row
```

Factor loadings were obtained.

```
cfa1.fig.b <- parameterEstimates(cfa1, standardized=TRUE) %>%
  filter(op == "=~") %>%
  select('Latent Factor'=lhs, #rename left hand side column
         Indicator=rhs, #rename right hand side column
         B=est,  #rename estimates
         SE=se,  #rename standard error
         Z=z,   # rename z statistics
         'p-value'=pvalue,  # rename p value
         Beta=std.all) %>%
  kable(digits = 3,   # rounded
        format="markdown",  #Latex markdown, show in rmd
        booktabs=TRUE, #Latex booktabs
        caption="Factor Loadings for 4 factor CFA") %>% #caption
  kable_styling(latex_options = "striped") %>% #alternate grayed rows
  row_spec(0, background = "#9999CC") #color first row
```

Variance were obtained.

```
#Some notes from tutorial: in the "Variances:" section, there is a dot before the observed variables na

type <- rep(c("Residual variance", "Total variance"), #generate a new column
            time = c(ncol(sc), 4))

variance <- parameterEstimates(cfa1, standardized=TRUE) %>% #obtain estimator
  filter(op == "~~") #select rows with variable "operator" = "~~"
variance <- variance[1:20,] #after 20th row, no variance any more. (covariance)
variance <- cbind(type, variance) #add column
cfa1.fig.c <- variance %>%select(Type = type, #select and rename
                   Indicator=rhs, #right high side column
                   B=est, #estimates
                   SE=se, #standard error
                   Z=z,  #z statistics
                   'p-value'=pvalue,# p value
                   Beta=std.all) %>%  # beta
  kable(digits = 3, #round
        format="markdown", #latex
        booktabs=TRUE, #latex
        caption="Variances for 4 factor CFA") %>% #caption
  kable_styling(latex_options = "striped") %>% #grayed every alternate row
  row_spec(0, background = "#9999CC") #color the first row
```

Covariance were obtained.

```
variance <- parameterEstimates(cfa1, standardized=TRUE) %>% #obtain estimates
  filter(op == "~~") #select rows
variance <- variance[21:26,] #select rows for covariance instead of variance
type <- paste(variance$lhs, "with", variance$rhs) #collapse columns
variance <- cbind(type, variance) #add columns
rownames(variance) <- NULL #remove row names
cfa1.fig.d <- variance %>%select(Type=type, #select and rename
                   B=est, #same idea with last section
                   SE=se,
                   Z=z,
```

```
                'p-value'=pvalue,
                Beta=std.all) %>%
  kable(digits = 3, #same idea with last section
        format="markdown",
        booktabs=TRUE,
        caption="Covariances for 4 factor CFA") %>%
  kable_styling(latex_options = "striped") %>%
  row_spec(0, background = "#9999CC")
```

```
cfa1.fig.a;cfa1.fig.b;cfa1.fig.c;cfa1.fig.d
```

Table 4: Goodness-of-fit and subjective indices of fit for 4 factor CFA

Measure	Value
chisq	159.112
df	98.000
pvalue	0.000
cfi	0.961
tli	0.953
rmsea	0.049
rmsea.pvalue	0.556
srmr	0.048

Table 5: Factor Loadings for 4 factor CFA

Latent Factor	Indicator	B	SE	Z	p-value	Beta
GSC	SDQ2N01	1.000	0.000	NA	NA	0.582
GSC	SDQ2N13	1.083	0.154	7.044	0	0.626
GSC	SDQ2N25	0.851	0.132	6.455	0	0.544
GSC	SDQ2N37	0.934	0.131	7.131	0	0.640
ASC	SDQ2N04	1.000	0.000	NA	NA	0.536
ASC	SDQ2N16	1.279	0.150	8.520	0	0.774
ASC	SDQ2N28	1.247	0.154	8.097	0	0.703
ASC	SDQ2N40	1.259	0.156	8.048	0	0.695
ESC	SDQ2N10	1.000	0.000	NA	NA	0.711
ESC	SDQ2N22	0.889	0.103	8.658	0	0.668
ESC	SDQ2N34	0.670	0.148	4.539	0	0.322
ESC	SDQ2N46	0.843	0.117	7.225	0	0.532
MSC	SDQ2N07	1.000	0.000	NA	NA	0.854
MSC	SDQ2N19	0.841	0.058	14.495	0	0.755
MSC	SDQ2N31	0.952	0.049	19.516	0	0.923
MSC	SDQ2N43	0.655	0.049	13.298	0	0.712

Table 6: Variances for 4 factor CFA

Type	Indicator	B	SE	Z	p-value	Beta
Residual variance	SDQ2N01	1.198	0.126	9.537	0	0.661
Residual variance	SDQ2N13	1.119	0.124	9.019	0	0.609

Type	Indicator	B	SE	Z	p-value	Beta
Residual variance	SDQ2N25	1.056	0.107	9.897	0	0.704
Residual variance	SDQ2N37	0.771	0.087	8.821	0	0.591
Residual variance	SDQ2N04	1.394	0.128	10.900	0	0.713
Residual variance	SDQ2N16	0.616	0.068	9.020	0	0.402
Residual variance	SDQ2N28	0.896	0.090	9.959	0	0.506
Residual variance	SDQ2N40	0.952	0.095	10.029	0	0.517
Residual variance	SDQ2N10	0.653	0.082	7.941	0	0.494
Residual variance	SDQ2N22	0.657	0.075	8.735	0	0.554
Residual variance	SDQ2N34	2.590	0.233	11.128	0	0.896
Residual variance	SDQ2N46	1.201	0.118	10.183	0	0.717
Residual variance	SDQ2N07	0.854	0.100	8.551	0	0.270
Residual variance	SDQ2N19	1.228	0.121	10.153	0	0.429
Residual variance	SDQ2N31	0.365	0.065	5.649	0	0.148
Residual variance	SDQ2N43	0.964	0.092	10.473	0	0.493
Total variance	GSC	0.613	0.137	4.464	0	1.000
Total variance	ASC	0.561	0.126	4.453	0	1.000
Total variance	ESC	0.668	0.116	5.749	0	1.000
Total variance	MSC	2.307	0.273	8.460	0	1.000

Table 7: Covariances for 4 factor CFA

Type	B	SE	Z	p-value	Beta
GSC with ASC	0.415	0.078	5.292	0.000	0.707
GSC with ESC	0.355	0.072	4.947	0.000	0.555
GSC with MSC	0.635	0.118	5.387	0.000	0.534
ASC with ESC	0.464	0.078	5.921	0.000	0.758
ASC with MSC	0.873	0.134	6.519	0.000	0.767
ESC with MSC	0.331	0.100	3.309	0.001	0.266

3.1.2 Model visualization

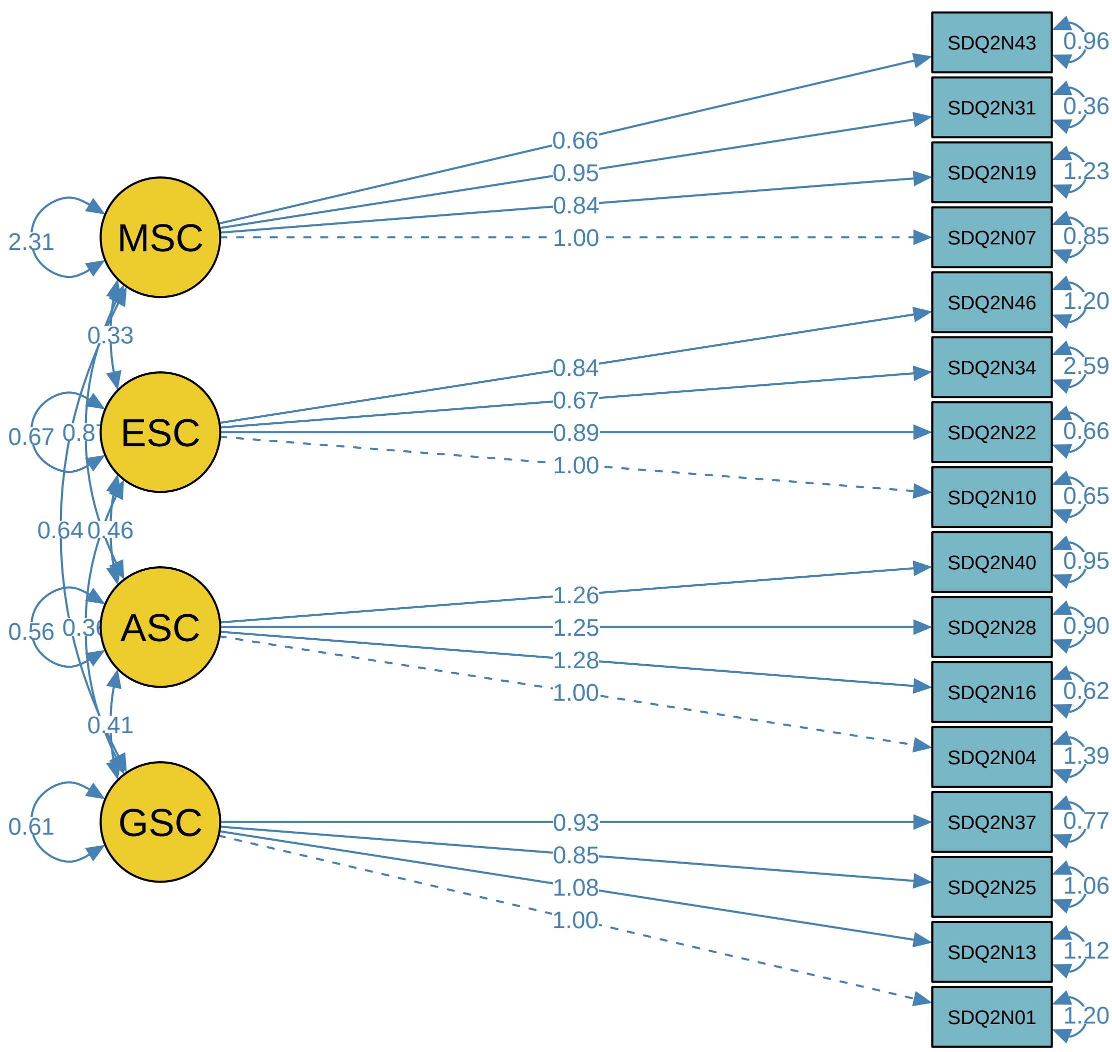
The 4 factor model was visualized.

```
library(semPlot)
library(wesanderson)#a handful of color palettes from Wes anderson movies
mycols <- wes_palette(name = "Zissou1", n = 4, type = "discrete")
colorlist <- list(man = mycols[2], lat = mycols[3])

semPaths(cfa1,
         "par",
         weighted = FALSE, #no weight
         curvature = 1,  #curvature strength
         shapeMan = "rectangle", #manifest variable's shape
         sizeMan = 8, # manifest variable's font size
         sizeMan2 = 4, # manifest variable's tile size
         rotation = 2, # turn vertical
         color = colorlist, #specify color by calling colorlist defined above
         edge.color = "steelblue", #specify line color
         edge.label.cex = 0.7,
```

```r
         title = T) #specify line label font size
title("Figure 4. Four factor self-concept CFA model diagram")
```

Figure 4. Four factor self−concept CFA model diagram



3.2 Alternative hypotheses 1: SC has two fctors (GSC, ASC)

3.2.1 Model estimation

Two factor reduced model was estimated herein.

```
#define model
model2 <- '# CFA model of self-concept (SC):
           GSC =~ SDQ2N01 + SDQ2N13 + SDQ2N25 + SDQ2N37
           ASC =~ SDQ2N04 + SDQ2N16 + SDQ2N28 + SDQ2N40 +
                  SDQ2N10 + SDQ2N22 + SDQ2N34 + SDQ2N46 +
                  SDQ2N07 + SDQ2N19 + SDQ2N31 + SDQ2N43
          '
# Estimate the model using the data defined earlier
cfa2 <- cfa(model2, data = sc)
```

Measures of goodness of fit and subjective indices of fit were obtained.

```
#turn off scientific notion
options(scipen = 999)

#obtain CFA measures
cfa2.measure <- fitMeasures(cfa2,    #obtain specified measured.
                            c("chisq",
                              "df",
                              "pvalue",
                              "cfi",
                              "tli",
                              "rmsea",
                              "rmsea.pvalue",
                              "srmr"))

#turn named vector to data frame
cfa2.fig.a <- cfa2.measure %>%
  tibble(name= names(cfa2.measure), value = cfa2.measure) %>% # vector to df
  select(Measure = name, Value = value) %>%  #select and rename columns
  mutate(Value = round(as.numeric(Value),3)) %>%  # round
  kable(format = "markdown",   # table aesthetics
        booktabs = T, #Latex booktabs
        caption =  #caption
          "Goodness-of-fit and subjective indices of fit for 2 factor CFA") %>%
  kable_styling(latex_options = "striped") %>% # gray every other row
  row_spec(0, background = "#9999CC") # color first row
```

Factor loadings were obtained.

```
cfa2.fig.b <- parameterEstimates(cfa2, standardized=TRUE) %>% # obtain estimates
  filter(op == "=~") %>%  #select "is measured by" rows
  select('Latent Factor'=lhs, #left hand side column
         Indicator=rhs, #right hand side column
         B=est, #estimates
         SE=se, #standard error
         Z=z, #z statistics
```

```
         'p-value'=pvalue, #p value
         Beta=std.all) %>%
  kable(digits = 3, #rounded to 3
        format="markdown", #Latex markdown
        booktabs=TRUE, #Latex booktabs
        caption="Factor Loadings for 2 factor CFA") %>% #caption
  kable_styling(latex_options = "striped") %>% #gray every other row
  row_spec(0, background = "#9999CC") # color the first row
```

Variance were obtained.

```
type <- rep(c("Residual variance", "Total variance"),
            time = c(ncol(sc), 2)) #create a new row clarifying types of variance

variance <- parameterEstimates(cfa2, standardized=TRUE) %>% #obtain estimates
  filter(op == "~~") #select "is correlated with" rows
variance <- variance[1:18,] #subset 1:18 rows (variance row)
variance <- cbind(type, variance) #add column
cfa2.fig.c <- variance %>%select(Type = type, #select and rename variables
                  Indicator=rhs, #right hand side column
                  B=est, #estimates
                  SE=se,#standard error
                  Z=z, #z statistics
                  'p-value'=pvalue, #p value
                  Beta=std.all) %>%
  kable(digits = 3, #rounded
        format="markdown",  #Latex markdown
        booktabs=TRUE, #Latex booktabs
        caption="Variances for 2 factor CFA") %>% #caption
  kable_styling(latex_options = "striped") %>% # gray every other row
  row_spec(0, background = "#9999CC") # color the variable row
```

Covariance were obtained.

```
variance <- parameterEstimates(cfa2, standardized=TRUE) %>%
  filter(op == "~~")
variance <- variance[19,]
type <- paste(variance$lhs, "with", variance$rhs)
variance <- cbind(type, variance)
rownames(variance) <- NULL
cfa2.fig.d <- variance %>%select(Type=type,
                  B=est,
                  SE=se,
                  Z=z,
                  'p-value'=pvalue,
                  Beta=std.all) %>%
  kable(digits = 3,
        format="markdown",
        booktabs=TRUE,
        caption="Covariances for 2 factor CFA") %>%
  kable_styling(latex_options = "striped") %>%
  row_spec(0, background = "#9999CC")
```

```
cfa2.fig.a;cfa2.fig.b;cfa2.fig.c;cfa2.fig.d
```

Table 8: Goodness-of-fit and subjective indices of fit for 2 factor CFA

Measure	Value
chisq	457.653
df	103.000
pvalue	0.000
cfi	0.776
tli	0.739
rmsea	0.114
rmsea.pvalue	0.000
srmr	0.101

Table 9: Factor Loadings for 2 factor CFA

Latent Factor	Indicator	B	SE	Z	p-value	Beta
GSC	SDQ2N01	1.000	0.000	NA	NA	0.595
GSC	SDQ2N13	1.048	0.151	6.930	0.000	0.619
GSC	SDQ2N25	0.860	0.131	6.542	0.000	0.562
GSC	SDQ2N37	0.890	0.128	6.957	0.000	0.623
ASC	SDQ2N04	1.000	0.000	NA	NA	0.485
ASC	SDQ2N16	1.263	0.170	7.440	0.000	0.692
ASC	SDQ2N28	1.276	0.177	7.221	0.000	0.651
ASC	SDQ2N40	1.235	0.176	7.026	0.000	0.618
ASC	SDQ2N10	0.581	0.123	4.736	0.000	0.343
ASC	SDQ2N22	0.558	0.117	4.786	0.000	0.348
ASC	SDQ2N34	0.065	0.161	0.406	0.685	0.026
ASC	SDQ2N46	0.514	0.132	3.885	0.000	0.270
ASC	SDQ2N07	2.069	0.262	7.885	0.000	0.790
ASC	SDQ2N19	1.871	0.242	7.721	0.000	0.751
ASC	SDQ2N31	2.021	0.247	8.192	0.000	0.875
ASC	SDQ2N43	1.442	0.193	7.481	0.000	0.700

Table 10: Variances for 2 factor CFA

Type	Indicator	B	SE	Z	p-value	Beta
Residual variance	SDQ2N01	1.170	0.127	9.216	0	0.646
Residual variance	SDQ2N13	1.134	0.127	8.906	0	0.617
Residual variance	SDQ2N25	1.026	0.107	9.582	0	0.684
Residual variance	SDQ2N37	0.799	0.090	8.842	0	0.612
Residual variance	SDQ2N04	1.495	0.134	11.171	0	0.764
Residual variance	SDQ2N16	0.799	0.076	10.490	0	0.521
Residual variance	SDQ2N28	1.018	0.095	10.695	0	0.576
Residual variance	SDQ2N40	1.138	0.105	10.828	0	0.618
Residual variance	SDQ2N10	1.166	0.103	11.364	0	0.882
Residual variance	SDQ2N22	1.043	0.092	11.360	0	0.879
Residual variance	SDQ2N34	2.888	0.251	11.510	0	0.999

Type	Indicator	B	SE	Z	p-value	Beta
Residual variance	SDQ2N46	1.554	0.136	11.425	0	0.927
Residual variance	SDQ2N07	1.191	0.123	9.654	0	0.377
Residual variance	SDQ2N19	1.247	0.124	10.067	0	0.436
Residual variance	SDQ2N31	0.575	0.073	7.852	0	0.234
Residual variance	SDQ2N43	0.996	0.095	10.442	0	0.510
Total variance	GSC	0.641	0.142	4.508	0	1.000
Total variance	ASC	0.461	0.114	4.034	0	1.000

Table 11: Covariances for 2 factor CFA

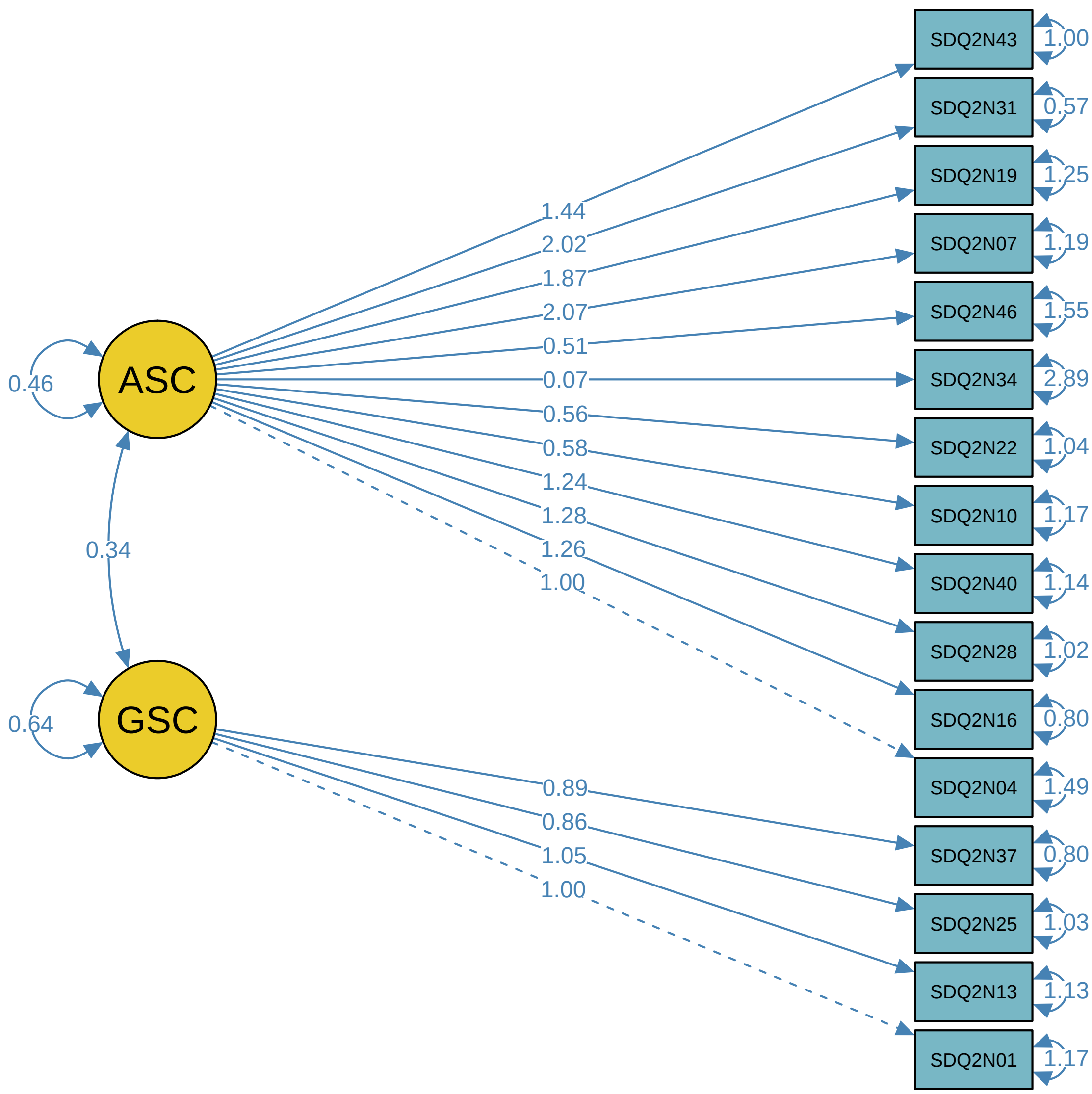
Type	B	SE	Z	p-value	Beta
GSC with ASC	0.34	0.068	4.975	0	0.626

3.2.2 Model visualization

Two factor model was visualized.

```
semPaths(cfa2,
         "par", #estimates showed
         weighted = FALSE,
         curvature = 1, #residual variance curvature
         shapeMan = "rectangle", #manifest variable shape
         sizeMan = 8, #manifest variable font size
         sizeMan2 = 4, #manifest variable square size
         rotation = 2, #literally
         color = colorlist, #color the shapes
         edge.color = "steelblue", #color edges
         edge.label.cex = 0.7) # edge label font size
title("Figure 5. Tow factor self-concept CFA model diagram")
```

Figure 5. Tow factor self−concept CFA model diagram



3.3 Alternative hypotheses 2: SC is unidimensional (only one SC factor)

3.3.1 Modeal estimation

The redued modeal was estimated.

```
#define model
model3 <- '# CFA model of self-concept (SC):
           SC =~ SDQ2N01 + SDQ2N13 + SDQ2N25 + SDQ2N37 +
                 SDQ2N04 + SDQ2N16 + SDQ2N28 + SDQ2N40 +
                 SDQ2N10 + SDQ2N22 + SDQ2N34 + SDQ2N46 +
                 SDQ2N07 + SDQ2N19 + SDQ2N31 + SDQ2N43
          '

# Estimate the model using the data defined earlier
cfa3 <- cfa(model3, data = sc)
```

Measures of goodness of fit and subjective indices of fit were obtained.

```
#turn off scientific notion
options(scipen = 999)

#obtain CFA measures
cfa3.measure <- fitMeasures(cfa3,    #obtain specified measured.
                            c("chisq",
                              "df",
                              "pvalue",
                              "cfi",
                              "tli",
                              "rmsea",
                              "rmsea.pvalue",
                              "srmr"))
#turn named vector to data frame
cfa3.fig.a <- cfa3.measure %>%
  tibble(name= names(cfa3.measure), value = cfa3.measure) %>% # vector to df
  select(Measure = name, Value = value) %>%  #select and rename columns
  mutate(Value = round(as.numeric(Value),3)) %>%  # round
  kable(format = "markdown",   # table aesthetics
        booktabs = T, #Latex booktabs
        caption =  #caption
          "Goodness-of-fit and subjective indices of fit for uni-factor CFA") %>%
  kable_styling(latex_options = "striped") %>% # gray every other row
  row_spec(0, background = "#9999CC") # color first row
```

Factor loadings were obtained.

```
cfa3.fig.b <- parameterEstimates(cfa3, standardized=TRUE) %>% # obtain estimates
  filter(op == "=~") %>%  #select "is measured by" rows
  select('Latent Factor'=lhs, #left hand side column
         Indicator=rhs, #right hand side column
         B=est, #estimates
         SE=se, #standard error
         Z=z, #z statistics
```

```
         'p-value'=pvalue, #p value
         Beta=std.all) %>%
  kable(digits = 3, #rounded to 3
        format="markdown", #Latex markdown
        booktabs=TRUE, #Latex booktabs
        caption="Factor Loadings for uni-factor CFA") %>% #caption
  kable_styling(latex_options = "striped") %>% #gray every other row
  row_spec(0, background = "#9999CC") # color the first row
```

Variance were obtained.

```
type <- rep(c("Residual variance", "Total variance"),
            time = c(ncol(sc), 1)) #create a new row clarifying types of variance

variance <- parameterEstimates(cfa3, standardized=TRUE) %>% #obtain estimates
  filter(op == "~~") #select "is correlated with" rows
variance <- variance[1:17,] #subset 1:18 rows (variance row)
variance <- cbind(type, variance) #add column
cfa3.fig.c <- variance %>%select(Type = type, #select and rename variables
                   Indicator=rhs, #right hand side column
                   B=est, #estimates
                   SE=se,#standard error
                   Z=z, #z statistics
                   'p-value'=pvalue, #p value
                   Beta=std.all) %>%
  kable(digits = 3, #rounded
        format="markdown",  #Latex markdown
        booktabs=TRUE, #Latex booktabs
        caption="Variances for 2 factor CFA") %>% #caption
  kable_styling(latex_options = "striped") %>% # gray every other row
  row_spec(0, background = "#9999CC") # color the variable row
```

```
cfa3.fig.a;cfa3.fig.b;cfa3.fig.c
```

Table 12: Goodness-of-fit and subjective indices of fit for uni-factor CFA

Measure	Value
chisq	531.918
df	104.000
pvalue	0.000
cfi	0.730
tli	0.688
rmsea	0.125
rmsea.pvalue	0.000
srmr	0.104

Table 13: Factor Loadings for uni-factor CFA

Latent Factor	Indicator	B	SE	Z	p-value	Beta
SC	SDQ2N01	1.000	0.000	NA	NA	0.368
SC	SDQ2N13	1.158	0.247	4.690	0.000	0.423
SC	SDQ2N25	0.903	0.209	4.330	0.000	0.366
SC	SDQ2N37	1.126	0.224	5.018	0.000	0.489
SC	SDQ2N04	1.407	0.278	5.063	0.000	0.499
SC	SDQ2N16	1.772	0.310	5.716	0.000	0.709
SC	SDQ2N28	1.775	0.317	5.605	0.000	0.662
SC	SDQ2N40	1.744	0.315	5.541	0.000	0.637
SC	SDQ2N10	0.859	0.197	4.362	0.000	0.370
SC	SDQ2N22	0.816	0.187	4.371	0.000	0.372
SC	SDQ2N34	0.181	0.222	0.815	0.415	0.053
SC	SDQ2N46	0.756	0.202	3.732	0.000	0.289
SC	SDQ2N07	2.743	0.471	5.826	0.000	0.765
SC	SDQ2N19	2.505	0.434	5.768	0.000	0.735
SC	SDQ2N31	2.711	0.454	5.970	0.000	0.857
SC	SDQ2N43	1.929	0.341	5.659	0.000	0.684

Table 14: Variances for 2 factor CFA

Type	Indicator	B	SE	Z	p-value	Beta
Residual variance	SDQ2N01	1.565	0.138	11.335	0.000	0.864
Residual variance	SDQ2N13	1.508	0.134	11.266	0.000	0.821
Residual variance	SDQ2N25	1.299	0.115	11.338	0.000	0.866
Residual variance	SDQ2N37	0.994	0.089	11.160	0.000	0.761
Residual variance	SDQ2N04	1.469	0.132	11.140	0.000	0.751
Residual variance	SDQ2N16	0.762	0.073	10.368	0.000	0.497
Residual variance	SDQ2N28	0.994	0.093	10.633	0.000	0.562
Residual variance	SDQ2N40	1.093	0.102	10.742	0.000	0.594
Residual variance	SDQ2N10	1.140	0.101	11.333	0.000	0.863
Residual variance	SDQ2N22	1.022	0.090	11.332	0.000	0.862
Residual variance	SDQ2N34	2.882	0.250	11.508	0.000	0.997
Residual variance	SDQ2N46	1.535	0.135	11.409	0.000	0.916
Residual variance	SDQ2N07	1.311	0.132	9.913	0.000	0.415
Residual variance	SDQ2N19	1.316	0.129	10.186	0.000	0.460
Residual variance	SDQ2N31	0.650	0.078	8.367	0.000	0.265
Residual variance	SDQ2N43	1.040	0.099	10.520	0.000	0.532
Total variance	SC	0.246	0.083	2.972	0.003	1.000

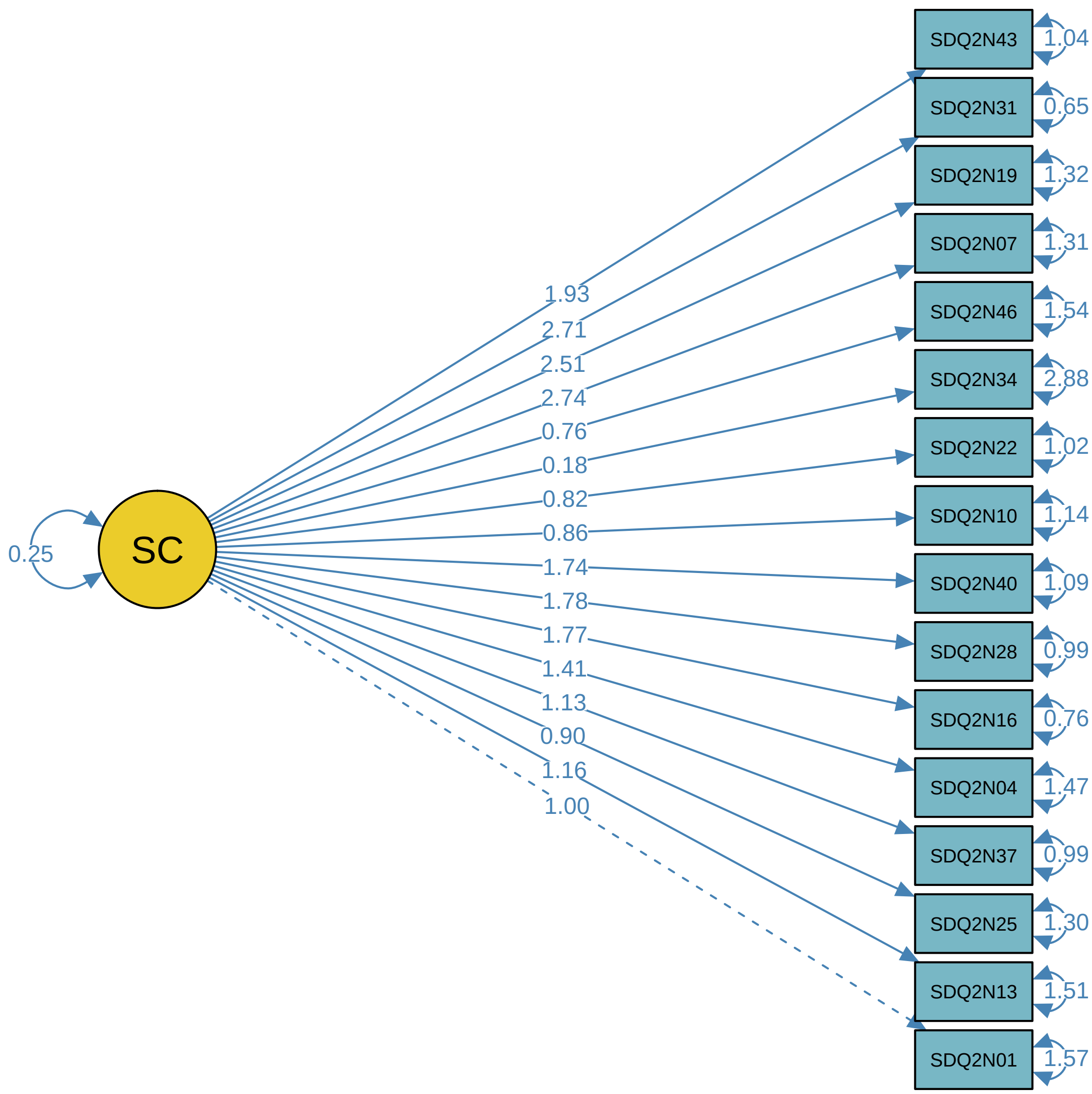
3.3.2 Model visualization

The uni-factor model was visualzied.

```
semPaths(cfa3,
         "par", #estimates
         weighted = FALSE,
         curvature = 1,
         shapeMan = "rectangle",#Man for manifest variable
         sizeMan = 8,
```

```
         sizeMan2 = 4, #color the squares
         rotation = 2, #turn vertical
         color = colorlist, #color the shapes
         edge.color = "steelblue", #color the edge
         edge.label.cex = 0.7)# resize the edge label font
title("Figure 6. Uni-factor self-concept CFA model diagram")
```

Figure 6. Uni−factor self−concept CFA model diagram



3.3 Model comparison

```
#combine measures from three models
comparison.tab <- rbind(cfa1.measure, cfa2.measure, cfa3.measure) %>% round(2)
#turn row name into a a new column (model)
model <- rownames(comparison.tab)
rownames(comparison.tab) <- NULL
comparison.tab <- cbind(model, comparison.tab) %>% data.frame()

turn.tab <- comparison.tab %>% t() %>% data.frame

#calculate model comparison indicators
turn.tab <- turn.tab %>%
  filter(!rownames(turn.tab) == "model")%>%
  mutate(X1 = as.numeric(X1),
         X2 = as.numeric(X2),
         X3 = as.numeric(X3)) %>%
  mutate(a = X2 - X1,
         b = X3 - X1,
         c = X3 - X2)

comparison.tab <- t(turn.tab) %>% data.frame

names <- c("Model1
           (4 factor)",
           "Model2
           (2 factor)",
           "Model3
           (uni-factor)",
           "Model Contrast: 1-2§",
           "Model Contrast: 1-3§",
           "Model Contrast: 2-3§")

#collapse columns for parsimony
comparison.tab <- comparison.tab %>%
  mutate(rmseap = paste(rmsea, "(", rmsea.pvalue,")"),
         chisq.df.p = paste(chisq, "(", df, ",", pvalue, ")"))

rownames(comparison.tab) <- names
model <- rownames(comparison.tab)
rownames(comparison.tab) <- NULL
comparison.tab <- cbind(model, comparison.tab) %>% data.frame()

#select and rename columns
comparison.tab <- comparison.tab %>%
  select(Model = model, 'Chi-square
                        (df, p)' = chisq.df.p,
         'CFI*' = cfi,
         'TLI†' = tli,
         'RMSEA
         (p)‡' = rmseap,
         'SRMR‡' = srmr)
```

Table 15: Model comparison

Model	Chi-square (df, p)	CFI*	TLI†	RMSEA (p)‡	SRMR‡
Model1 (4 factor)	159.11 (98 , 0)	0.96	0.95	0.05 (0.56)	0.05
Model2 (2 factor)	457.65 (103 , 0)	0.78	0.74	0.11 (0)	0.10
Model3 (uni-factor)	531.92 (104 , 0)	0.73	0.69	0.12 (0)	0.10
Model Contrast: 1-2§	298.54 (5 , 0)	-0.18	-0.21	0.06 (-0.56)	0.05
Model Contrast: 1-3§	372.81 (6 , 0)	-0.23	-0.26	0.07 (-0.56)	0.05
Model Contrast: 2-3§	74.27 (1 , 0)	-0.05	-0.05	0.01 (0)	0.00

* CFI>0.95 indicates well fitting

† TLI close to 1 indicates well fitting

‡ <0.05 indicates good fit

§ Constrast of models rather than direct model values

```r
#display the table
comparison.tab %>%
  kable(booktabs = T,
        linesep = "",
        align = c("l",rep("r",5)),
        caption = "Model comparison",
        ) %>%
  kable_styling(full_width = T) %>%
  column_spec(1, width = "3.5cm") %>%
  column_spec(2, width = "3cm") %>%
  column_spec(5, width = "2.5cm") %>%
  footnote(symbol = c("CFI>0.95 indicates well fitting",
                      "TLI close to 1 indicates well fitting",
                      "<0.05 indicates good fit",
                      "Constrast of models rather than direct model values")) %>%
  row_spec(4:6, background = "grey")
```

Chi-square Test of Model fit is a traditional likelihood ratio test statistics. It tests the null hypothesis that the model is adequate. chi^2 is sensitive to sample size and non-normal data. For each of the three models, the null hypothesis was rejected, indicating the model was not adequate. However, according to the normality test, our data did not follow normal distribution, which might distort the results. Moreover, it is still important in model comparisons. Smaller chi-square values reflect that the estimated model is able to adequately reproduce the observed sample statistics whereas larger values reflect that some aspect of the hypothesized model is inconsistent with characteristics of the observed sample. Four-factor model had chi^2 value much smaller than the other two models, indicating it fitted better.

CFI is an incremental index measuring the proportionate improvement in fit with nested models. CFI > 0.95 indicates well-fitting and larger CFI means better fit. Our four factor model had a CFI=0.96, while CFIs of other two model all fell below 0.95 by around 0.20.

TLI is quite similar as CFI but non-normed. TLI closer to 1 indicates well-fitting. The four-, two- and uni-factor had TLI 0.95, 0.74 and 0.69, respectively. No doubt, four factor model is most close to 1.

RMSEA is an absolute index. It describes how well the model fits the data. RMSEA<0.05 indicate good fit. Four factor model had a RMSEA = 0.05 and RMSEA of the other two models had values more than doubled it. Four-factor model fitted the data better again with regard to this indicator.

SRMR represents the average residual value of the fit. In well-fitting models, it will be small, less than 0.05. Our nested four-factor model had an SRMR of 0.05, much smaller than those of the other two, indicating well fit.

In summary, the four-factor model meet the standard of good fit in all indexes including CFI, TLI, RMSE and SRMR. And compared to other two models, it had values indicating better fit. We should adopt this model before better solution has been found.