R Workshop: Factor Analysis

Brier Gallihugh, M.S.

2023-06-18

Table of contents

Creating Data
EFA Assumptions
EFA Factor Structure
EFA Factor Structure Assumptions
Calculating Reliability

Creating Data

- (1) Create a data set using the bfi dataset in the psych package
- (2) Create a correlation matrix of the bfi items using the cor() function
- (3) Create an APA Style correlation output within Word
- (4) Round correlation matrix to 2 decimal places for readability in R

Means, standard deviations, and correlations with confidence intervals

```
11. E1 -0.02 0.38 .42 -.71** -.75** -.61*
[-.11, .77] [-.89, -.30] [-.91, -.40] [-.85, -.13]
```

```
[-.80, .03] [-.91, -.38] [-.92, -.41] [-.89, -.30] [.63, .95]
                                                                .39
-.79**
            -.32
                         -.20
                                      -.12
                                                   .39
[-.93, -.46] [-.71, .23] [-.65, .35] [-.60, .42] [-.15, .75] [-.15, .75]
-.81**
            -.40
                         -.28
                                      -.20
                                                   .49
                                                                .54*
[-.93, -.50] [-.76, .14] [-.70, .27] [-.65, .35] [-.03, .80] [.04, .83]
.83**
             .30
                         .21
                                      .10
                                                   -.41
                                                                -.51
                        [-.34, .65] [-.44, .58] [-.76, .13] [-.81, .01]
[.55, .94]
            [-.25, .70]
.84**
            .30
                                                   -.41
                                                                -.49
                         . 18
                                      .11
                         [-.37, .63] [-.43, .59] [-.76, .12] [-.80, .03]
            [-.25, .70]
[.58, .95]
            .52*
                                     .47
.59*
                          .55*
                                                   -.70**
                                                                -.69**
            [.01, .81]
                         [.06, .83] [-.06, .79] [-.89, -.30] [-.89, -.27]
[.10, .84]
11
            12
                         13
                                    14
```

Note. M and SD are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates p < .05. ** indicates p < .01.

```
Α1
          A2
                АЗ
                     A4
                           A5
                                C1
                                      C2
                                           C3
                                                C4
                                                      C5
                                                           E1
                                                                 E2
A1 1.00 -0.36 -0.31 -0.19 -0.23
                              0.00 -0.02 -0.08 0.12 0.10 0.13
A2 -0.36 1.00 0.51
                   0.35
                         0.40
                              0.14  0.20  0.19  -0.20  -0.16  -0.26  -0.26
A3 -0.31 0.51 1.00 0.33
                         0.59
                              A4 -0.19 0.35 0.33 1.00
                         0.35
                              0.10 0.22 0.06 -0.16 -0.23 -0.16 -0.19
A5 -0.23 0.40 0.59 0.35
                         1.00
                              0.09 0.07 0.09 -0.14 -0.13 -0.29 -0.31
C1 0.00 0.14 0.08 0.10
                         0.09
                              1.00 0.44 0.32 -0.38 -0.32 -0.09 -0.13
C2 -0.02 0.20 0.11 0.22
                         0.07
                              0.44
                                   1.00 0.41 -0.42 -0.31 0.02 -0.02
C3 -0.08 0.19 0.10 0.06 0.09 0.32 0.41 1.00 -0.35 -0.30 0.04 0.01
C4 0.12 -0.20 -0.09 -0.16 -0.14 -0.38 -0.42 -0.35
                                              1.00 0.52 0.16 0.22
C5 0.10 -0.16 -0.14 -0.23 -0.13 -0.32 -0.31 -0.30 0.52 1.00 0.08 0.26
E1 0.13 -0.26 -0.28 -0.16 -0.29 -0.09 0.02 0.04 0.16 0.08 1.00 0.54
E2 0.14 -0.26 -0.29 -0.19 -0.31 -0.13 -0.02 0.01 0.22 0.26 0.54 1.00
E3 -0.08 0.32 0.44 0.26 0.47 0.14 0.12 0.02 -0.06 -0.17 -0.38 -0.43
E4 -0.09 0.27 0.37 0.30 0.50 0.13 0.06 0.04 -0.10 -0.16 -0.44 -0.55
E5 0.00 0.34 0.25 0.23 0.24 0.21 0.31 0.25 -0.28 -0.24 -0.30 -0.36
```

```
E3
           E4
                 E5
A1 -0.08 -0.09
               0.00
A2 0.32 0.27
               0.34
A3 0.44 0.37
               0.25
A4 0.26 0.30
              0.23
A5 0.47 0.50
              0.24
C1 0.14 0.13 0.21
C2 0.12 0.06 0.31
C3 0.02 0.04 0.25
C4 -0.06 -0.10 -0.28
C5 -0.17 -0.16 -0.24
E1 -0.38 -0.44 -0.30
E2 -0.43 -0.55 -0.36
E3 1.00 0.47 0.36
E4 0.47 1.00 0.26
E5 0.36 0.26 1.00
```

EFA Assumptions

```
#Barlett Test for New Scale
cortest.bartlett(cor_proposed_scale, n = 500)

#KMO for New Scale
KMO(cor_proposed_scale)

#Determinent for New Scale
det(cor_proposed_scale)

3
```

- 1 Run a Bartlett test on the correlation matrix. Ideally, this should have a p value of less than .05
- (2) Run a KMO on the proposed correlation matrix. Ideally this is greater than KMO = .90
- (3) Find the determinant of the correlation matrix. This should be less than .00001

\$chisq [1] 2225.86 \$p.value [1] 0 \$df

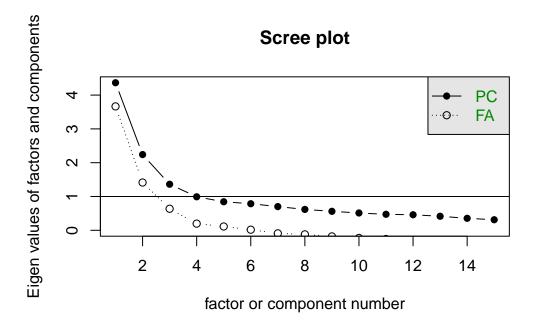
[1] 105

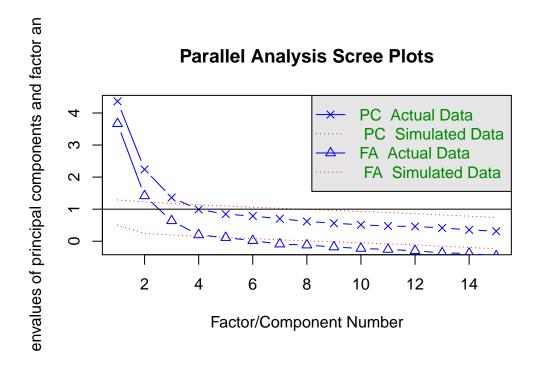
Kaiser-Meyer-Olkin factor adequacy Call: KMO(r = cor_proposed_scale) Overall MSA = 0.82MSA for each item = A1 A2 ΑЗ C1 C2 СЗ C4 C5 E1 E2 E3 E4 **A4** A5 0.73 0.86 0.85 0.86 0.83 0.83 0.77 0.79 0.76 0.78 0.85 0.80 0.88 0.84 0.85 [1] 0.0109611

EFA Factor Structure

psych::scree(cor_proposed_scale)

- (3) Run an orthogonal rotation factor analysis using the fa() function
- (4) Print the output fit measures using the print.psych() function. The SORT = TRUE argument sorts the factor loading by loading magnitude.
- (5) Run an oblique rotation factor analysis using the fa() function
- (6) Print the output again using the print.psych() function





```
# Suggests 4 Factor Solution
  # Orthogonal (Non Correlated)
  orthoFA3 <- fa(r = cor_proposed_scale, nfactors = 4,rotate = 'varimax', use = "pairwise.co
  #Show All Info
  print.psych(orthoFA3, sort = TRUE)
                                                                         4
  # Oblique (Correlated)
  obliqueFA3 <- fa(r = cor_proposed_scale, nfactors = 4,rotate = 'oblimin', use = "pairwise.
  print.psych(obliqueFA3, sort = TRUE)
                                                                         6
Parallel analysis suggests that the number of factors = 4 and the number of components = 3
Factor Analysis using method = minres
Call: fa(r = cor_proposed_scale, nfactors = 4, rotate = "varimax",
    use = "pairwise.complete.obs")
Standardized loadings (pattern matrix) based upon correlation matrix
   item
          MR1
                MR2
                     MR3
                            MR4
                                  h2
                                       u2 com
A3
      3 0.73 0.05 0.17 -0.19 0.60 0.40 1.3
      5 0.67 0.05 0.27 -0.07 0.53 0.47 1.4
A5
A2
        0.57 0.21 0.13 -0.28 0.47 0.53 1.9
```

0.53 0.07 0.45 0.16 0.51 0.49 2.2

4 0.43 0.19 0.13 -0.11 0.25 0.75 1.7

E3

Α4

```
C2
    7 0.17 0.70 -0.10 0.12 0.54 0.46 1.2
    9 0.03 -0.69 -0.18 0.23 0.57 0.43 1.4
C4
    6 0.07 0.56 0.08 0.08 0.34 0.66 1.1
C1
C5
    C3
    8 0.11 0.55 -0.10 -0.03 0.33 0.67 1.2
E5
    15 0.28 0.38 0.32 0.10 0.33 0.67 3.0
E2
    11 -0.22 -0.01 -0.60 0.07 0.42 0.58 1.3
E1
E4
   14 0.42 0.05 0.57 0.08 0.51 0.49 1.9
    1 -0.30 -0.02 -0.02 0.56 0.40 0.60 1.5
A 1
```

MR1 MR2 MR3 MR4
SS loadings 2.24 2.15 1.91 0.60
Proportion Var 0.15 0.14 0.13 0.04
Cumulative Var 0.15 0.29 0.42 0.46
Proportion Explained 0.32 0.31 0.28 0.09
Cumulative Proportion 0.32 0.64 0.91 1.00

Mean item complexity = 1.6
Test of the hypothesis that 4 factors are sufficient.

df null model = 105 with the objective function = 4.51 df of the model are 51 and the objective function was 0.3

The root mean square of the residuals (RMSR) is 0.03The df corrected root mean square of the residuals is 0.04

Fit based upon off diagonal values = 0.99 Measures of factor score adequacy

Call: fa(r = cor_proposed_scale, nfactors = 4, rotate = "oblimin",
 use = "pairwise.complete.obs")

Standardized loadings (pattern matrix) based upon correlation matrix

E3 13 0.39 0.01 -0.34 0.30 0.51 0.49 2.9

C2 7 0.10 0.71 0.18 0.17 0.54 0.46 1.3 9 0.08 -0.68 0.19 0.21 0.57 0.43 1.4 C4 6 -0.03 0.57 -0.05 0.11 0.34 0.66 1.1 C1 C3 8 0.09 0.56 0.16 0.00 0.33 0.67 1.2 10 0.01 -0.55 0.18 0.11 0.38 0.62 1.3 C5 E5 15 0.15 0.35 -0.26 0.18 0.33 0.67 2.8 12 0.02 -0.03 0.86 0.04 0.73 0.27 1.0 E2 11 -0.11 0.05 0.60 -0.01 0.42 0.58 1.1 E1 E4 14 0.28 -0.01 -0.50 0.20 0.51 0.49 1.9 1 -0.46 0.03 -0.03 0.48 0.40 0.60 2.0 A 1

MR1 MR2 MR3 MR4
SS loadings 2.33 2.15 1.89 0.53
Proportion Var 0.16 0.14 0.13 0.04
Cumulative Var 0.16 0.30 0.42 0.46
Proportion Explained 0.34 0.31 0.27 0.08
Cumulative Proportion 0.34 0.65 0.92 1.00

With factor correlations of MR1 MR2 MR3 MR4 MR1 1.00 0.21 -0.44 0.07 MR2 0.21 1.00 -0.17 -0.02 MR3 -0.44 -0.17 1.00 -0.08 MR4 0.07 -0.02 -0.08 1.00

Mean item complexity = 1.5
Test of the hypothesis that 4 factors are sufficient.

df null model = 105 with the objective function = 4.51 df of the model are 51 and the objective function was 0.3

The root mean square of the residuals (RMSR) is 0.03 The df corrected root mean square of the residuals is 0.04

Fit based upon off diagonal values = 0.99 Measures of factor score adequacy

MR1 MR2 MR3 MR4 Correlation of (regression) scores with factors 0.91 0.89 0.91 0.70 Multiple R square of scores with factors 0.82 0.80 0.83 0.49 Minimum correlation of possible factor scores 0.64 0.60 0.65 -0.02

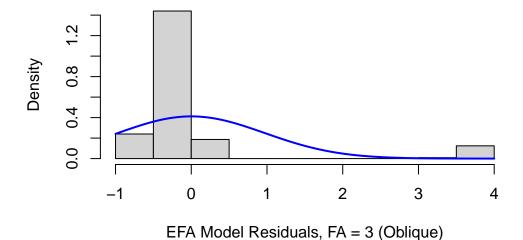
Tip

More often than not, an oblique rotation will be the best fit for your data as it assumes that your items are correlated with one another

EFA Factor Structure Assumptions

```
#Standard Residuals
obliqueFA3Residuals <- scale(obliqueFA3$residual)
#Test Normality
shapiro.test(obliqueFA3Residuals)
#Histogram
hist(obliqueFA3Residuals, col = 'lightgrey',
    main="", xlab = "EFA Model Residuals, FA = 3 (Oblique)",
    probability = TRUE)
curve(dnorm(x, mean = mean(obliqueFA3Residuals),
    sd = sd(obliqueFA3Residuals)),
    add = TRUE, lwd = 2, col = 'blue')</pre>
```

- (1) Assess the residuals of your desired factor loading solution using the scale() function in combination with extracting the residuals using object\$residuals notation.
- (2) Statistical test of the factor solution residuals using the shapiro.test() function.
- (3) Graphical depiction of the solution residuals with a normal curve overlay in the color blue



Shapiro-Wilk normality test

```
data: obliqueFA3Residuals
W = 0.46308, p-value < 2.2e-16</pre>
```

Calculating Reliability

```
#Items
Factor1<- c("A1","A2","A3","A4","A5")
                                                                            1
Factor2<- c("C1", "C2", "C3", "C4", "C5")
                                                                            2
Factor3<- c("E1","E2","E3","E4","E5")
                                                                            (3)
Overall <- c("A1", "A2", "A3", "A4", "A5", "C1", "C2", "C3", "C4", "C5", "E1", "E2", "E3", "E4", "E5") (4)
#Reliability Factor 1
psych::alpha(proposed_scale[,Factor1], check.keys = TRUE)
                                                                            (5)
#Reliability Factor 2
psych::alpha(proposed_scale[, Factor2], check.keys = TRUE)
                                                                            (6)
#Reliablity Factor 3
psych::alpha(proposed_scale[, Factor3], check.keys = TRUE)
                                                                            (7)
#Overall Reliability
psych::alpha(proposed_scale[, Overall], check.keys = TRUE)
                                                                            (8)
```

- (1) Create a subset of items to represent Factor 1
- (2) Create a subset of items to represent Factor 2
- (3) Create a subset of items to represent Factor 3
- (4) Create a subset of items to represent Overall
- (5) Determine the reliability of Factor 1 using the alpha() function in the psych package. check.keys ensures that items that load negatively are reverse coded.
- 6 Determine the reliability of Factor 2 using the alpha() function in the psych package. check.keys ensures that items that load negatively are reverse coded.
- (7) Determine the reliability of Factor 3 using the alpha() function in the psych package. check.keys ensures that items that load negatively are reverse coded.
- (8) Determine the reliability of Overall using the alpha() function in the psych package. check.keys ensures that items that load negatively are reverse coded.

Feldt 0.69 0.73 0.77 Duhachek 0.69 0.73 0.77

Reliability if an item is dropped:

raw alpha std.alpha G6(smc) average r S/N alpha se var.r med.r A1-0.74 0.75 0.70 0.42 2.9 0.019 0.011 0.38 A2 0.66 0.67 0.63 0.33 2.0 0.025 0.019 0.32 A3 0.64 0.65 0.59 0.31 1.8 0.027 0.007 0.35 Α4 0.72 0.73 0.69 0.40 2.7 0.021 0.017 0.38 0.66 0.34 2.1 0.025 0.010 0.34 A5 0.68 0.62

Item statistics

n raw.r std.r r.cor r.drop mean sd
A1- 500 0.61 0.60 0.42 0.36 4.6 1.4
A2 500 0.73 0.75 0.66 0.57 4.8 1.1
A3 500 0.77 0.78 0.74 0.61 4.6 1.3
A4 500 0.66 0.64 0.48 0.41 4.7 1.5
A5 500 0.73 0.73 0.66 0.55 4.5 1.3

Non missing response frequency for each item

1 2 3 4 5 6 miss
A1 0.33 0.28 0.14 0.13 0.10 0.02 0
A2 0.01 0.05 0.05 0.19 0.40 0.30 0
A3 0.03 0.06 0.08 0.19 0.36 0.27 0
A4 0.05 0.07 0.07 0.16 0.25 0.41 0
A5 0.03 0.06 0.09 0.23 0.33 0.26 0

Reliability analysis

Call: psych::alpha(x = proposed_scale[, Factor2], check.keys = TRUE)

raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r 0.75 0.75 0.72 0.38 3 0.018 4.2 0.94 0.36

95% confidence boundaries

lower alpha upper

Feldt 0.71 0.75 0.78 Duhachek 0.71 0.75 0.78

Reliability if an item is dropped:

raw_alpha std.alpha G6(smc) average_r S/N alpha se var.r med.r C1 0.71 0.71 0.67 0.38 2.5 0.021 0.0064 0.38 C2 0.69 0.70 0.64 0.36 2.3 0.022 0.0061 0.33 СЗ 0.72 0.73 0.68 0.40 2.6 0.020 0.0060 0.40

```
C4- 0.67 0.68 0.63 0.35 2.2 0.024 0.0035 0.32 C5- 0.72 0.72 0.66 0.39 2.5 0.021 0.0021 0.39
```

Item statistics

n raw.r std.r r.cor r.drop mean sd C1 500 0.67 0.70 0.57 0.49 4.5 1.2 C2 500 0.72 0.73 0.63 0.54 4.3 1.3 C3 500 0.66 0.67 0.54 0.46 4.3 1.3 C4- 500 0.75 0.75 0.67 0.59 4.4 1.3 C5- 500 0.73 0.69 0.58 0.50 3.6 1.6

Non missing response frequency for each item

1 2 3 4 5 6 miss
C1 0.02 0.04 0.11 0.22 0.39 0.21 0
C2 0.03 0.09 0.10 0.25 0.34 0.18 0
C3 0.03 0.08 0.11 0.27 0.35 0.16 0
C4 0.24 0.32 0.18 0.18 0.07 0.02 0
C5 0.15 0.19 0.13 0.24 0.14 0.13 0

Reliability analysis

Call: psych::alpha(x = proposed_scale[, Factor3], check.keys = TRUE)

raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r 0.78 0.78 0.75 0.41 3.5 0.015 4.2 1.1 0.41

95% confidence boundaries

lower alpha upper

Feldt 0.75 0.78 0.81 Duhachek 0.75 0.78 0.81

Reliability if an item is dropped:

raw_alpha std.alpha G6(smc) average_r S/N alpha se var.r med.r E1-0.74 0.73 0.69 0.41 2.8 0.019 0.0102 0.40 E2-0.70 0.70 0.65 0.37 2.3 0.022 0.0063 0.37 E3 0.74 0.73 0.69 0.41 2.8 0.018 0.0144 0.40 0.40 2.6 0.020 0.0065 0.37 0.73 E4 0.72 0.67 E5 0.78 0.78 0.73 0.47 3.5 0.016 0.0043 0.46

Item statistics

n raw.r std.r r.cor r.drop mean sd E1- 500 0.75 0.73 0.63 0.56 4.0 1.6 E2- 500 0.81 0.79 0.74 0.65 3.8 1.6 E3 500 0.72 0.73 0.62 0.55 4.1 1.4

```
E4 500 0.75 0.75 0.67 0.59 4.4 1.5
E5 500 0.60 0.63 0.47 0.41 4.5 1.3
```

Non missing response frequency for each item

1 2 3 4 5 6 miss
E1 0.23 0.23 0.15 0.18 0.11 0.10 0
E2 0.19 0.23 0.12 0.22 0.14 0.09 0
E3 0.06 0.10 0.13 0.28 0.28 0.15 0
E4 0.06 0.10 0.09 0.15 0.34 0.26 0
E5 0.04 0.06 0.10 0.24 0.32 0.24 0

Reliability analysis

Call: psych::alpha(x = proposed_scale[, Overall], check.keys = TRUE)

raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r 0.82 0.82 0.85 0.23 4.5 0.012 4.3 0.73 0.23

95% confidence boundaries

lower alpha upper

Feldt 0.79 0.82 0.84 Duhachek 0.79 0.82 0.84

Reliability if an item is dropped:

	raw_alpha	${\tt std.alpha}$	G6(smc)	$average_r$	S/N	alpha se	var.r	med.r
A1-	0.82	0.82	0.85	0.25	4.5	0.012	0.021	0.25
A2	0.80	0.80	0.84	0.22	4.0	0.013	0.023	0.22
A3	0.80	0.80	0.83	0.22	4.0	0.013	0.020	0.22
A4	0.81	0.81	0.84	0.23	4.2	0.013	0.023	0.23
A5	0.80	0.80	0.83	0.22	4.0	0.013	0.020	0.22
C1	0.81	0.81	0.85	0.24	4.4	0.012	0.022	0.24
C2	0.81	0.81	0.84	0.24	4.3	0.012	0.021	0.23
СЗ	0.81	0.82	0.85	0.24	4.5	0.012	0.020	0.23
C4-	0.81	0.81	0.84	0.23	4.2	0.013	0.022	0.23
C5-	0.81	0.81	0.84	0.23	4.2	0.013	0.023	0.23
E1-	0.81	0.81	0.84	0.23	4.2	0.013	0.020	0.23
E2-	0.80	0.80	0.83	0.22	4.1	0.013	0.020	0.22
E3	0.80	0.80	0.84	0.22	4.1	0.013	0.020	0.22
E4	0.80	0.80	0.84	0.22	4.0	0.013	0.020	0.22
E5	0.80	0.80	0.84	0.23	4.1	0.013	0.023	0.19

Item statistics

n raw.r std.r r.cor r.drop mean sd A1- 500 0.36 0.36 0.28 0.24 4.6 1.4

```
A2
    500
         0.60
               0.62
                      0.59
                              0.53
                                    4.8 1.1
    500
         0.61
                0.61
                      0.59
                                    4.6 1.3
ΑЗ
                              0.52
A4
    500
         0.52
                0.52
                      0.46
                              0.42
                                    4.7 1.5
A5
    500
         0.61
                0.62
                      0.60
                              0.53
                                    4.5 1.3
    500
         0.43
                0.45
                      0.39
                                    4.5 1.2
C1
                              0.33
C2
    500
         0.44
                0.46
                                    4.3 1.3
                      0.42
                              0.34
C3
    500
         0.37
                0.40
                      0.33
                              0.27
                                     4.3 1.3
                0.52
C4- 500
         0.52
                      0.49
                              0.42
                                    4.4 1.3
C5- 500
         0.53
                0.52
                      0.47
                                    3.6 1.6
                              0.41
E1- 500
         0.54
                0.51
                      0.46
                              0.42
                                    4.0 1.6
E2- 500
         0.62
                0.59
                      0.57
                              0.52
                                    3.8 1.6
E3
    500
         0.60
                0.59
                      0.56
                                    4.1 1.4
                              0.51
    500
E4
         0.61
                0.60
                      0.57
                              0.52
                                    4.4 1.5
E5
    500
         0.58
                0.58
                      0.54
                              0.49
                                    4.5 1.3
```

Non missing response frequency for each item

```
1
                3
                     4
                           5
                                6 miss
A1 0.33 0.28 0.14 0.13 0.10 0.02
                                     0
A2 0.01 0.05 0.05 0.19 0.40 0.30
                                     0
A3 0.03 0.06 0.08 0.19 0.36 0.27
                                     0
A4 0.05 0.07 0.07 0.16 0.25 0.41
                                     0
A5 0.03 0.06 0.09 0.23 0.33 0.26
                                     0
C1 0.02 0.04 0.11 0.22 0.39 0.21
                                     0
C2 0.03 0.09 0.10 0.25 0.34 0.18
                                     0
C3 0.03 0.08 0.11 0.27 0.35 0.16
                                     0
C4 0.24 0.32 0.18 0.18 0.07 0.02
                                     0
C5 0.15 0.19 0.13 0.24 0.14 0.13
                                     0
E1 0.23 0.23 0.15 0.18 0.11 0.10
                                     0
E2 0.19 0.23 0.12 0.22 0.14 0.09
                                     0
E3 0.06 0.10 0.13 0.28 0.28 0.15
                                     0
E4 0.06 0.10 0.09 0.15 0.34 0.26
                                     0
E5 0.04 0.06 0.10 0.24 0.32 0.24
                                     0
```



If you have more than one factor, your scale is no longer one (or uni) dimensional. As such, the idea of an "overall" reliability is questionable at best. Further, all reliability estimates are sample dependent. For non-sample dependent metrics, one should consider Item Response Theory (IRT)