Psychology 405: Psychometric Theory
Scale Construction

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

May, 2025

Outline

Scale construction: A 10 steps program

1. Personality scales are not created in a theoretical vacuum. Perhaps the most important step in developing a new scale is a consideration of what is the construct of interest. What is it, what are manifestations of it, what is it not, and what should it not relate to.
2. Then, what is the population of interest? Are they young or old, highly literate, or somewhat challenged by literacy. Write items suitable for the population of interest.
3. Give the items to the participants. Make sure that they are engaged in the task.

Scale construction: A 10 steps program (continued)

4. To analyze the data, it is necessary to enter the data into a machine readable form.
   - This is a source of error. Double check for data entry errors.
   - Double entry (two different people enter the data and then the two files are automatically compared) is recommended.
   - Even better is automatic data entry (but then you need to check and double check the program).
   - my.data <- read.file() #go find the file on your computer
   - my.data <- read.file(myfile) #if you have the file name some
   - my.data <- read.clipboard() #if you have already copied the data to the clipboard
5. Run basic descriptive statistics to do one more check for errors. Graphically check as well.
   - describe(my.data)
   - pairs.panels(my.data)
6. Form the variance/covariance matrix from the items and examine the dimensionality of the resulting space.

Scale construction: A 10 steps program (continued)

7. Apply various data reduction techniques (factor analysis, principal components analysis, cluster analysis).

`fa`	For most factor analysis and rotation/transformation algorithms (need to specify the number of factors)
`irt.fa`	If you have polytomous or dichotomous items and want to take an IRT approach.
`principal`	aka `pca` for principal components analysis
`fa.parallel`	For parallel analysis and scree tests
`vss`	The Very Simple Structure criterion as well as the Minimum Average Parcel test
`nfactors`	Combine a number of different tests
`iclust`	Cluster analysis of items shows structure pretty well.
`bestScales`	Apply an empirical scale construction procedure.

Scale construction: A 10 steps program (continued)

8. Form composite scales of the selected items. Check these scales for various measures of internal consistency.
   - Form a list of items to score. Can use `make.keys` if desired.
   - `scoreItems` wlll give convetional reliability statistics and raw scores. If given correlation matrix input, will return reliability statistics.
   - `scoreOverlap`: Given a correlation matrix, will find scale statistics correcting for item overlap.
   - `scoreFast` and `scoreVeryFast` will give just scores given data and a keys.list (Meant for large data sets)
   - `bestScales` (For empirical scale construction)
   - `alpha` – just one scale at a time – Do you really want to do that?
   - `testRetest` Score identical items for two different times, organize the output to reflect variance components.

Scale construction: A 10 steps program (continued)

9. Validity
   - Discriminant validity requires that the scales not correlate with other, unrelated traits.
   - Convergent validity requires that the scale do correlate with other, alternative measures of the same trait.
   - Scale discriminant and convergent validity at the item level is reported in `scoreOverlap`.
   - Predictive and concurrent validity require alternative measures of the same and different constructs.
10. Cross validation to show the results are not sample dependent.
   - Validate on a new sample
   - Cross vaildate on a holdout sample
   - KFold cross validation
   - Bootstrap validation

Basic item development

As a demonstration of scale construction and validation, consider the following problem. N self report items are given to a number of people. This inventory is composed of subsets of items that are believed to measure different traits. In addition, each subject is rated by a friend on those same traits. There are several questions we can ask of these data:

1. Do the items form reliable scales?
2. What are the correlations of these scales?
3. Do the scales correlate with the peer ratings?
4. Can we empirically find a better structure of the items?
5. Do these revised scales show greater independence, reliability, and validity?

Item writing

To show the procedures, 12 students in a personality research course spent several weeks learning about each of four personality dimensions. Each student then wrote five items to assess each of four constructs.

1. Need for Achievement
2. Anxiety
3. Sociability
4. Impulsivity

As a group they examined all of the items and formed the best 80 items into one questionnaire with 20 items believed to measure each of the constructs. An additional four items were the simple stem: “I think I am ... ”.They administered this questionnaire to approximately ten friends each whom they also rated on these four constructs. Thus, we have a data set of about 110 participants assessed on 89 items (the 84 self report items and the 4 peer ratings + Gender).

These four sets of items can be seen as samples from four domains.

Initial data reading

The data, item labels, and scoring keys are saved on a web server. They may be accessed by the read.table(file.name) command. We then use the dim command to find out the dimensions of the data file as well as the names command to find out what the names are.

R code

```
prq.data.name <-
    "https://personality-project.org/courses/405/prq.data.csv"
prq.dictionary <-
   "https://personality-project.org/courses/405/prq.dictionary.csv"
prq.data <- read.file(prq.data.name)
prq.dictionary <- read.file(prq.dictionary)
dim(prq.data)
names(prq.data)
```

```
> dim(prq)
[1] 75  91
names(prq)
  names(prq)
 [1] "Exp"         "Subject"     "NeedAch"     "Anxiety"     "Sociability" "Impulsivity"
 [7] "Gender"      "q1"          "q2"          "q3"          "q4"          "q5"
[13] "q6"          "q7"          "q8"          "q9"          "q10"         "q11"
[19] "q12"         "q13"         "q14"         "q15"         "q16"         "q17"
[25] "q18"         "q19"         "q20"         "q21"         "q22"         "q23"
[31] "q24"         "q25"         "q26"         "q27"         "q28"         "q29"
[37] "q30"         "q31"         "q32"         "q33"         "q34"         "q35"
[43] "q36"         "q37"         "q38"         "q39"         "q40"         "q41"
```

What is the structure of a dictionary file?

R code

```
dim(prq.dictionary)
colnames(prq.dictionary]
headTail(prq.dictionary}
```

```
> dim(prq.dictionary)
[1] 89  2
> colnames(prq.dictionary)
[1] "Item"    "Content"
> headTail(prq.dictionary)
          Item                                                                Content
1       NeedAch                                                              NeedAch
2       Anxiety                                                              Anxiety
3   Sociability                                                          Sociability
4   Impulsivity                                                          Impulsivity
...         <NA>                                                                 <NA>
86          q81  I believe that if something is worth doing, it is worth doing well
87          q82                                  I am more emotional than my friends
88          q83                                      I am a very sociable person
89          q84                                         I am an impulsive person
```

Data checking

Always check the data first. Use the `describe` function. / /

R code

```
describe(prq)
```

```
  describe(prq)
           vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
Exp           1 75 5.17 2.64      6    5.21 2.97   1   9     8 -0.19    -1.28 0.31
Subject       2 75 4.85 2.68      5    4.75 2.97   1  10     9  0.24    -1.05 0.31
NeedAch       3 75 6.39 1.92      7    6.48 1.48   2  10     8 -0.40    -0.64 0.22
Anxiety       4 75 5.24 2.28      5    5.21 2.97   1  10     9  0.09    -1.18 0.26
Sociability   5 75 6.15 2.13      7    6.31 1.48   1   9     8 -0.69    -0.60 0.25
Impulsivity   6 75 5.16 2.35      5    5.20 2.97   1   9     8 -0.13    -1.32 0.27
Gender        7 74 1.51 0.50      2    1.52 0.00   1   2     1 -0.05    -2.02 0.06
q1            8 75 4.27 1.15      4    4.34 1.48   1   6     5 -0.52    -0.08 0.13
q2            9 75 3.37 1.39      3    3.33 1.48   1   6     5  0.21    -0.73 0.16
q3           10 75 4.36 1.34      5    4.48 1.48   1   6     5 -0.57    -0.51 0.15
q4           11 75 4.04 1.33      4    4.08 1.48   1   6     5 -0.11    -0.75 0.15
q5           12 75 4.35 1.16      5    4.44 1.48   1   6     5 -0.74     0.03 0.13
q6           13 75 3.21 1.41      3    3.13 1.48   1   6     5  0.43    -0.85 0.16
q7           14 75 4.17 1.54      5    4.28 1.48   1   6     5 -0.44    -0.99 0.18.
...
q80          87 75 3.91 1.30      4    3.92 1.48   1   6     5 -0.16    -0.73 0.15
q81          88 75 4.32 1.22      4    4.43 1.48   1   6     5 -0.57     0.13 0.14
q82          89 75 3.84 1.46      4    3.92 1.48   1   6     5 -0.39    -0.74 0.17
q83          90 75 4.08 1.33      4    4.10 1.48   2   6     4 -0.35    -1.06 0.15
q84          91 75 3.89 1.33      4    3.92 1.48   1   6     5 -0.32    -0.80 0.15
```

Data checking

In doing this, we discovered (on the first pass through the data) that one of the variables had a range of 46 rather than the 6 that was appropriate. Correcting the data, we can start over again. Even with well meaning, careful data entry, mistakes will happen in data entry. It is recommended that data be entered twice and then compared using software that compares the two files line by line and entry by entry. In all cases, make sure to describe the data and check that the ranges are appropriate for the data.
Thus, the data were edited and the prior steps were done again until there were no incorrectly entered subjects. One error that makes data checking complicated is a blank field in Excel is read improperly. However, if we copy the data file to the clipboard and then use the `read.clipboard.tab` function, this solves that problem. Note that the describe output shows that some variables do not have as many subjects as others.

Score the scales

1. Forming scale scores as linear sums (or averages) of the items is easy to do in R.
2. One technique (not recommended) is to do a series of recodings, creating new variables for each scale.
3. A simpler technique, using the `scoreItems` function from the *psych* package does this for all scales defined in a matrix of keys (the keys matrix).
4. This is essentially a matrix of -1, 0, and 1s where 0 means don't include the item in the scale, and a 1 means to include it. -1 means to reverse key the item.
5. If the data set has column names, the keys can just be a list of item names (with - preceding items to reverse) (This is the preferred way).

Making up the scoring keys

R code

```
nach  <- cs(q1,q5,q9,q13,q17, q21,q25,q29,q33,q37,q41,q45,q49,
 -q53,q57,q61, -q65, -q69, q73, q77, q81)

anx <- cs( q2, q6,q10, q14,  -q18, q22,q26, q30, q34, q38, q42, q46,
 q50, q54, q58,q62, q66,-q70, q74, q78, q82)

soc <-  cs(q3, q7, -q11, q15,q19, q23,q27, q31, q35, q39, q43, -q47,
 -q51, q55, q59, -q63, q67,  -q71, q75, -q79, q83)

imp <- cs( -q4,q8, -q12, q16,-q20, q24, q28, q32, -q36, q40, q44,
 q48, q52,q56, -q60, -q64,-q68,   -q72, q76,q80,q84)

prq.keys <- list(nach=nach,anxiety = anx, sociability=soc,
  impulsivity=imp,
 PeerNach= "NeedAch", PeerAnx = "Anxiety",PeerSoc = "Sociability",
       PeerImp="Impulsivity", gender="Gender")
```

The prq keys

```
 prq.keys
$nach
 [1] "q1"   "q5"   "q9"   "q13"  "q17"  "q21"  "q25"  "q29"  "q33"  "q37"  "q41"  "q45"
[13] "q49"  "-q53" "q57"  "q61"  "-q65" "-q69" "q73"  "q77"  "q81"

$anxiety
 [1] "q2"   "q6"   "q10"  "q14"  "-q18" "q22"  "q26"  "q30"  "q34"  "q38"  "q42"  "q46"
[13] "q50"  "q54"  "q58"  "q62"  "q66"  "-q70" "q74"  "q78"  "q82"

$sociability
 [1] "q3"   "q7"   "-q11" "q15"  "q19"  "q23"  "q27"  "q31"  "q35"  "q39"  "q43"  "-q47"
[13] "-q51" "q55"  "q59"  "-q63" "q67"  "-q71" "q75"  "-q79" "q83"

$impulsivity
 [1] "-q4"  "q8"   "-q12" "q16"  "-q20" "q24"  "q28"  "q32"  "-q36" "q40"  "q44"  "q48"
[13] "q52"  "q56"  "-q60" "-q64" "-q68" "-q72" "q76"  "q80"  "q84"

$PeerNach
[1] "NeedAch"

$PeerAnx
[1] "Anxiety"

$PeerSoc
[1] "Sociability"

$PeerImp
[1] "Impulsivity"

$gender
[1] "Gender"
```

An example of a keys list for the bfi

R code

```
bfi.keys <-
  list(agree=c("-A1","A2","A3","A4","A5"),
  conscientious=c("C1","C2","C3","-C4","-C5"),
  extraversion=c("-E1","-E2","E3","E4","E5"),
  neuroticism=c("N1","N2","N3","N4","N5"),
  openness = c("O1","-O2","O3","O4","-O5"))
  #or use the cs function
bfikeys <- list(agree=cs(-A1,A2,A3,A4,A5), con =cs(C1,C2,C3,-C4,-C5),
extra=cs(-E1,-E2,E3,E4,E5),
 N=cs(N1,N2,N3,N4,N5),Open = cs(O1,-O2,O3,O4,-O5))
bfi.keys #show them
```

```
bfi.keys
$agree
[1] "-A1" "A2"  "A3"  "A4"  "A5"

$conscientious
[1] "C1"  "C2"  "C3"  "-C4" "-C5"

$extraversion
[1] "-E1" "-E2" "E3"  "E4"  "E5"

$neuroticism
[1] "N1" "N2" "N3" "N4" "N5"

$openness
[1] "O1"  "-O2" "O3"  "O4"  "-O5"
```

Score the items

We use the `scoreItems` function.

We first do this just for the items. The item.scores is a list of multiple values:

1. scores – the actual scores for each subject
2. missing – where there any missing values for any subject?
3. alpha – coefficient alpha for each scale
4. av.r – the average r within each scale
5. n.items – how many items in each scale?
6. item.cor – the correlation of each item with each scale
7. cor – the correlation matrix of the scales (based upon the correlations of the items - with SAPA data this will differ from correlating the scales)
8. corrected – the raw correlations of the scales (below the diagonal), the alpha reliabilities (on the diagonal), and the intercorrelations corrected for unreliability (above the diagonal).

Using scoreItems

R code

```
> prq.scores <- scoreItems(prq.keys,prq)
> prq.scores
```

```
Call: scoreItems(keys = prq.keys, items = prq)

(Unstandardized) Alpha:
      nach anxiety sociability impulsivity PeerNach PeerAnx PeerSoc PeerImp gender
alpha 0.87    0.85        0.87        0.87        1       1       1       1      1

Standard errors of unstandardized Alpha:
      nach anxiety sociability impulsivity PeerNach PeerAnx PeerSoc PeerImp gender
ASE   0.03   0.032        0.03        0.03      NaN     NaN     NaN     NaN    NaN

Average item correlation:
          nach anxiety sociability impulsivity PeerNach PeerAnx PeerSoc PeerImp gender
average.r 0.24    0.21        0.24        0.23      NaN     NaN     NaN     NaN    NaN

Median item correlation:
       nach     anxiety sociability impulsivity    PeerNach     PeerAnx     PeerSoc     PeerIm
       0.25        0.23        0.28        0.25          NA          NA          NA          N

 Guttman 6* reliability:
         nach anxiety sociability impulsivity PeerNach PeerAnx PeerSoc PeerImp gender
Lambda.6 0.98    0.97        0.98        0.98     0.93    0.88     0.9    0.86   0.88

Signal/Noise based upon av.r :
             nach anxiety sociability impulsivity PeerNach PeerAnx PeerSoc PeerImp gender
Signal/Noise  6.5     5.7         6.5         6.4      NaN     NaN     NaN     NaN    NaN
```

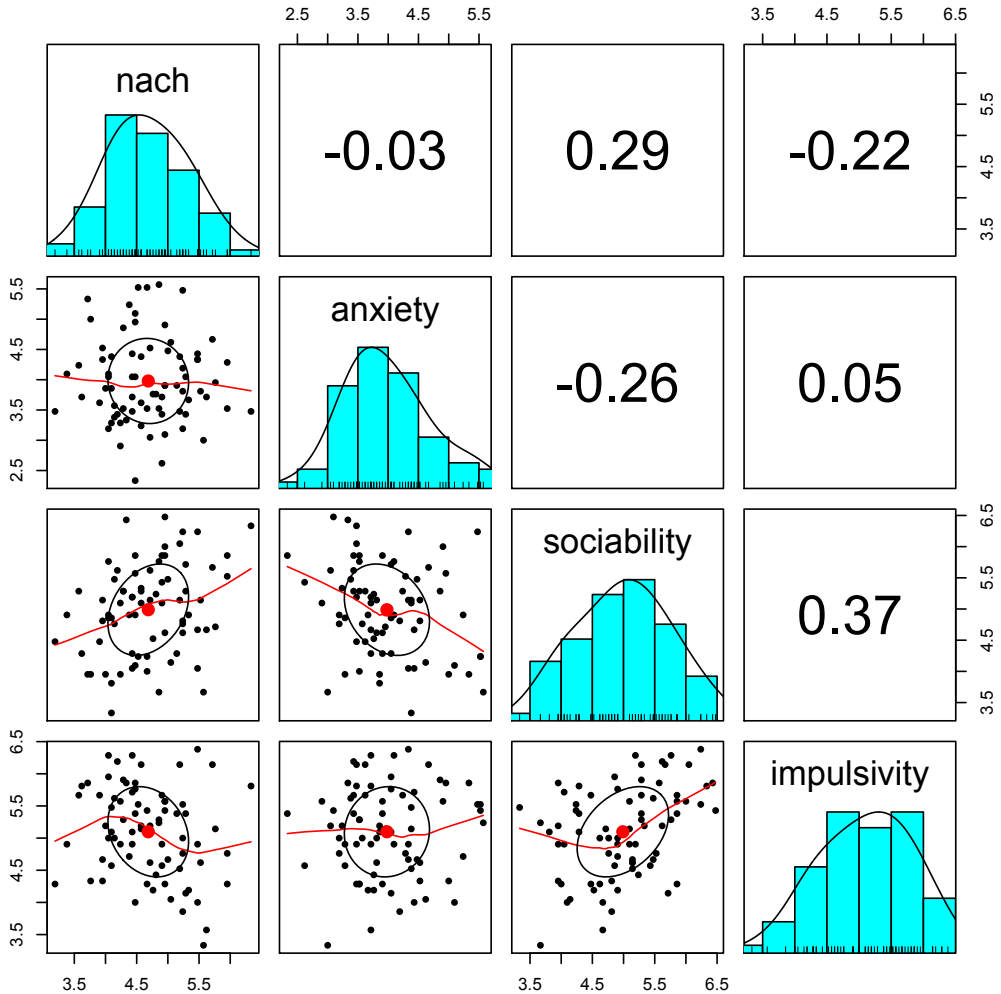
Show more of the output

scores

```
Scale intercorrelations corrected for attenuation
 raw correlations below the diagonal, alpha on the diagonal
 corrected correlations above the diagonal:
             nach anxiety sociability impulsivity PeerNach PeerAnx PeerSoc PeerImp  gender
nach        0.867  -0.038      0.3400      -0.251   0.2434  -0.039  -0.181  -0.277 -0.0862
anxiety    -0.033   0.852     -0.2979       0.058   0.1245   0.643  -0.195   0.136  0.1933
sociability 0.295  -0.256      0.8662       0.428   0.0039  -0.177   0.586   0.242  0.0663
impulsivity -0.218  0.049      0.3703       0.865  -0.3109   0.059   0.372   0.535  0.0791
PeerNach    0.227   0.115      0.0036      -0.289   1.0000   0.207  -0.077  -0.304 -0.0011
PeerAnx    -0.036   0.594     -0.1644       0.055   0.2068   1.000  -0.102  -0.030  0.3733
PeerSoc    -0.169  -0.180      0.5455       0.346  -0.0767  -0.102   1.000   0.293  0.0919
PeerImp    -0.258   0.126      0.2256       0.498  -0.3041  -0.030   0.293   1.000  0.0545
gender     -0.080   0.178      0.0617       0.074  -0.0011   0.373   0.092   0.054  1.0000

 In order to see the item by scale loadings and frequency counts of the data
 print with the short option = FALSE
```

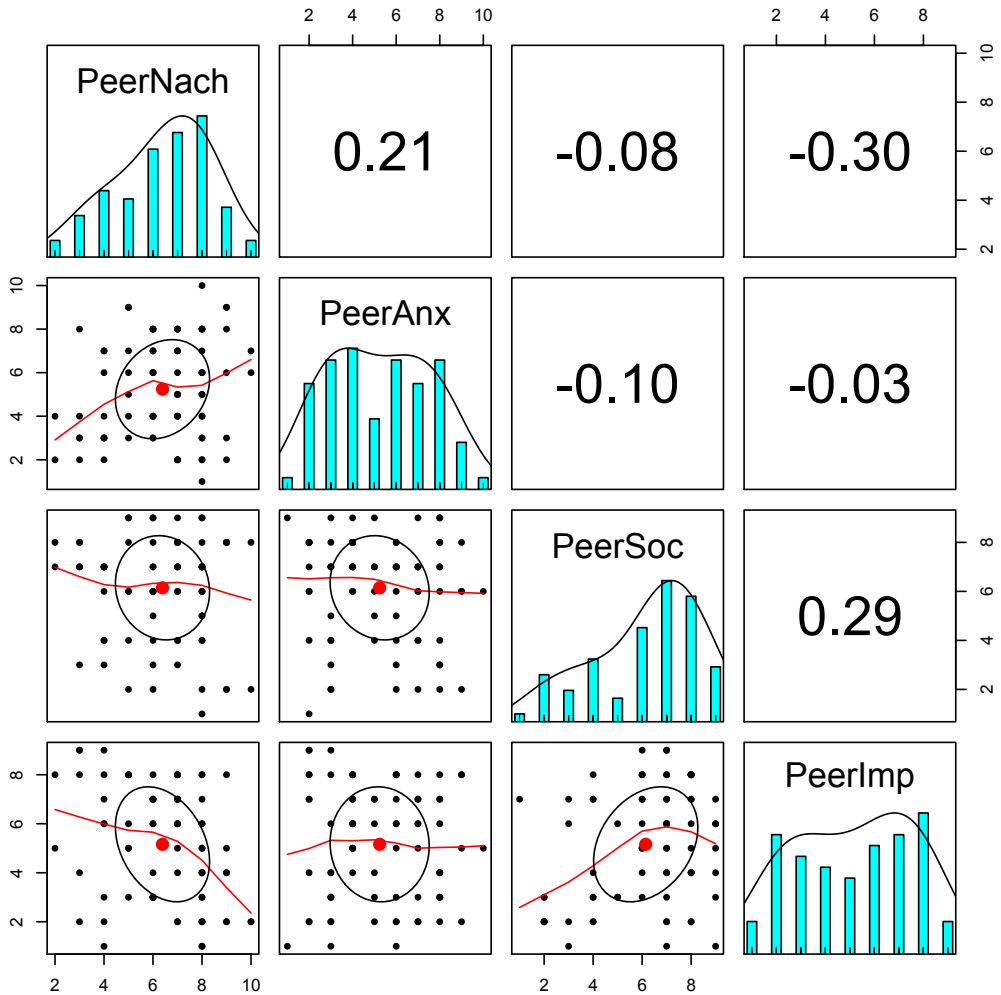
Display the four self report dimensions

pairs.panels(prq.scores$scores[,1:4]) # note that scores is an object in prq.scores



Show the peer rating structure

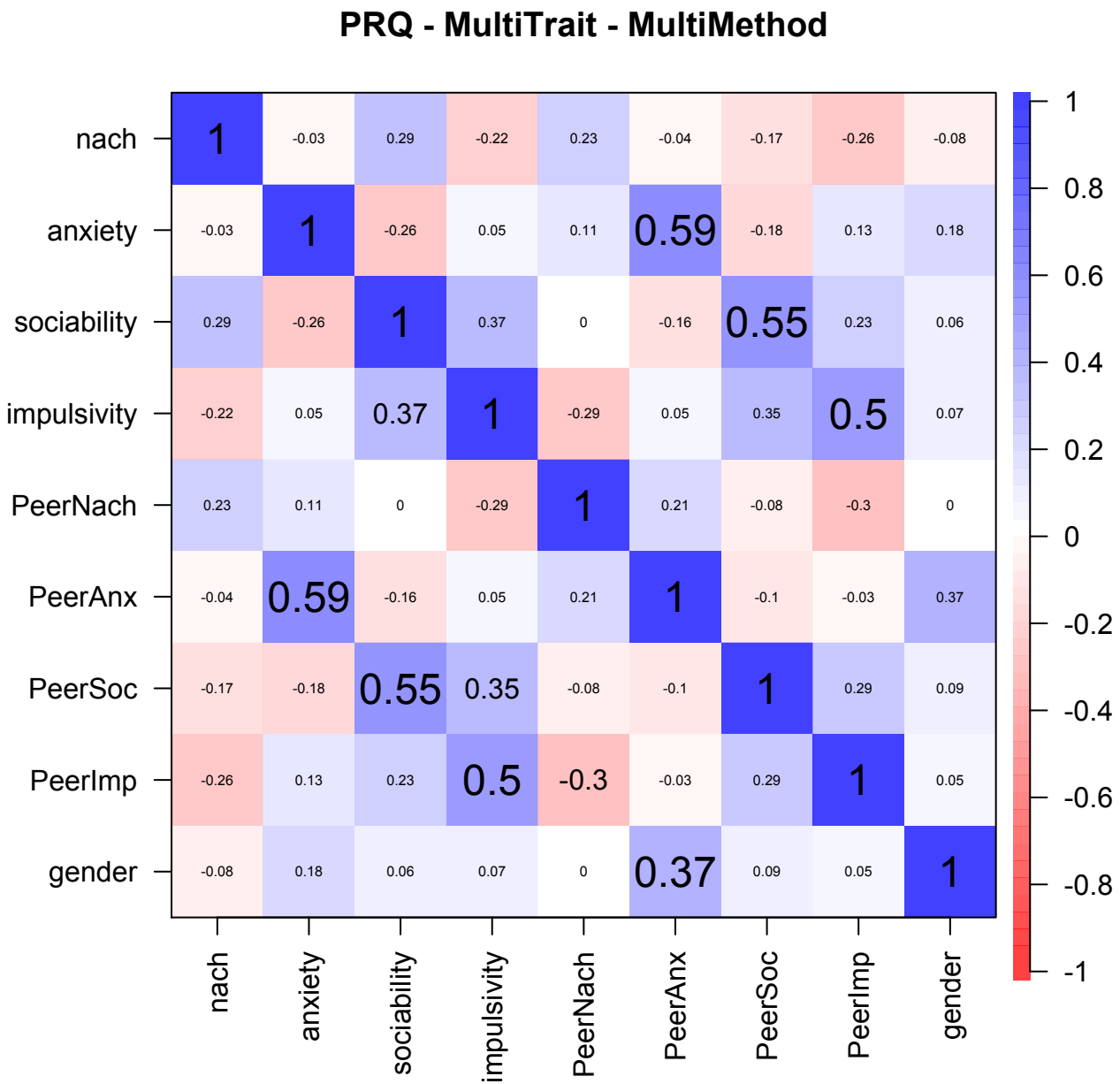
pairs.panels(prq.scores$scores[,5:8])



The Multi-Trait- Multi- Method Matrix

1. Correlations within method combine trait and method variance
    - What is the structure of NASI within self report
    - What is the structure of NASI within peer report
2. Correlations across method show trait variance
    - Do the self report dimensions match the peer ratings?
    - Note the correlations of gender differ between self and peer report. What could account for this difference?

Show the MMTM matrix graphically – corPlot(prq.scores$scores)



Factor Analysis

The items analysed were meant to represent four constructs. Given the previous analysis, they probably do. But what if we did not know how many separate dimensions were in the data? Is it possible to find out? Three alternative procedure address this question.

1. Principal components analysis
2. Factor analysis
3. Cluster analysis

All three of these procedures are attempting to approximate the nvar * nvar correlation matrix R with a matrix of lesser rank, one that is nvar * nf. That is, can we find a Factor (Component or Cluster) such that

$$R \approx FF' + U^2 \tag{1}$$

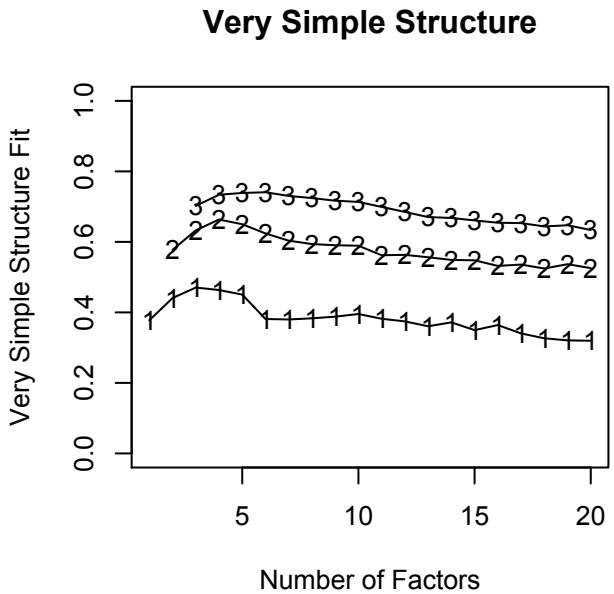
or

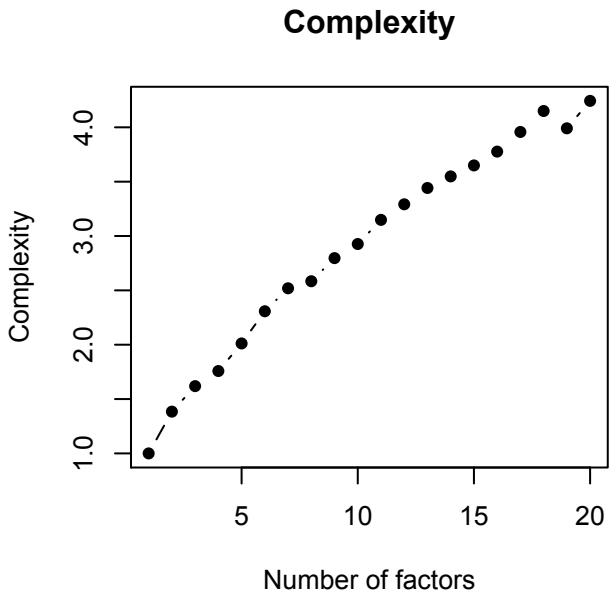
$$R \approx CC' \tag{2}$$

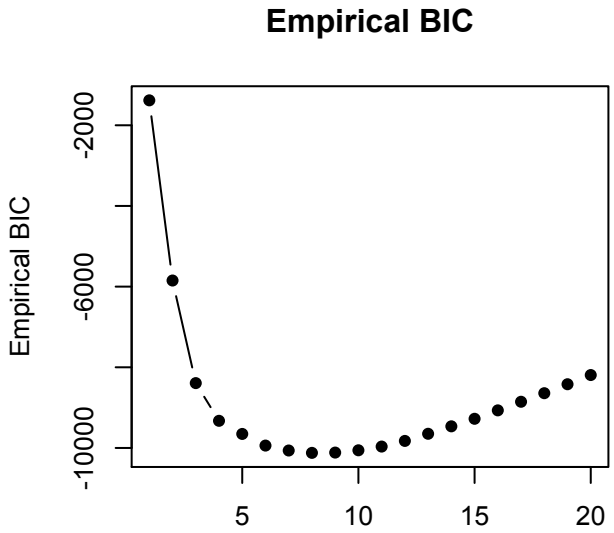
Factor analysis of PRQ

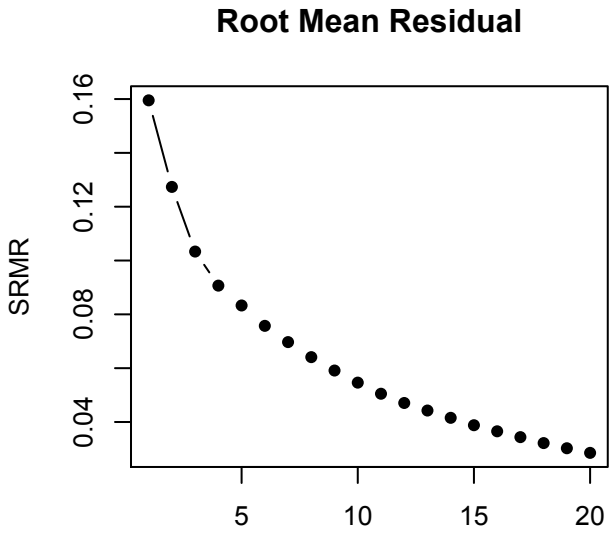
1. We normally need more people than items to make the matrix invertible (for MLE)
2. Can be solved in either case by using minimum residuals (OLS)
3. Can be solved by the `fa` function using minres option
4. How many factors to extract is a perpetual problem.
   - nfactors(prq)
   - Use VSS 3 (complexity 1) or 4 (complexity 2)
   - Use MAPS 12
   - Empirical BIC 8 factors
5. Theory says 4

VSS of prq


Very Simple Structure
Very Simple Structure Fit
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
Number of Factors


Complexity
Complexity
1.0
2.0
3.0
4.0
5
10
15
20
Number of factors


Empirical BIC
Empirical BIC
-10000
-6000
-2000
5
10
15
20


Root Mean Residual
SRMR
0.04
0.08
0.12
0.16
5
10
15
20

The number of factors problem

R code

```
nfactors(prq[8:91])
```

```
Number of factors
Call: vss(x = x, n = n, rotate = rotate, diagonal = diagonal, fm = fm,
    n.obs = n.obs, plot = FALSE, title = title, use = use, cor = cor)
VSS complexity 1 achieves a maximum of 0.47  with  3  factors
VSS complexity 2 achieves a maximum of 0.66  with  4  factors
The Velicer MAP achieves a minimum of 0.02  with  12  factors
Empirical BIC achieves a minimum of  -10121.68  with  8  factors
Sample Size adjusted BIC achieves a minimum of  5408.41  with  20  factors

Statistics by number of factors
   vss1 vss2   map  dof chisq prob sqresid  fit RMSEA   BIC SABIC complex eChisq  SRMR eCRMS
1  0.38 0.00 0.034 3402 13272    0     242 0.38  0.20 -1416  9306     1.0  13306 0.160 0.161
2  0.44 0.58 0.028 3319 12738    0     164 0.58  0.19 -1591  8869     1.4   8481 0.127 0.131
3  0.47 0.63 0.024 3237 12287    0     116 0.70  0.19 -1689  8513     1.6   5583 0.103 0.107
4  0.46 0.66 0.023 3156 11904    0      93 0.76  0.19 -1722  8225     1.8   4299 0.091 0.095
5  0.45 0.65 0.022 3076 11593    0      80 0.79  0.19 -1688  8007     2.0   3628 0.083 0.089
6  0.38 0.62 0.022 2997 11273    0      68 0.83  0.19 -1667  7779     2.3   2999 0.076 0.082
7  0.38 0.60 0.022 2919 10970    0      59 0.85  0.19 -1633  7567     2.5   2540 0.070 0.076
8  0.38 0.59 0.021 2842 10675    0      51 0.87  0.19 -1596  7362     2.6   2149 0.064 0.071
9  0.39 0.59 0.021 2766 10400    0      44 0.89  0.19 -1542  7176     2.8   1828 0.059 0.066
10 0.40 0.59 0.021 2691 10128    0      39 0.90  0.19 -1490  6991     2.9   1561 0.055 0.062
11 0.38 0.56 0.021 2617  9856    0      34 0.91  0.19 -1443  6805     3.1   1333 0.050 0.058
12 0.37 0.56 0.021 2544  9596    0      30 0.92  0.19 -1388  6630     3.3   1158 0.047 0.055
13 0.36 0.56 0.021 2472  9350    0      27 0.93  0.19 -1323  6468     3.4   1024 0.044 0.053
14 0.37 0.55 0.021 2401  9106    0      24 0.94  0.19 -1261  6307     3.5    903 0.042 0.050
15 0.35 0.55 0.021 2331  8854    0      22 0.94  0.19 -1210  6137     3.6    787 0.039 0.047
16 0.36 0.53 0.021 2262  8627    0      19 0.95  0.19 -1139  5990     3.8    698 0.037 0.045
17 0.34 0.54 0.022 2194  8402    0      18 0.95  0.19 -1071  5844     4.0    618 0.034 0.043
18 0.33 0.52 0.022 2127  8170    0      16 0.96  0.19 -1013  5691     4.2    541 0.032 0.041
```

Find a 4 factor as well as a 4 component solution – very similar

```
f4 <- fa(prq[8:91], 4)
p4 <- principal(prq[8:91],4)
> factor.congruence(f4,p4)
```

```
factor.congruence ( f4 , p4 )
      RC1   RC2   RC4   RC3
MR1  0.99  0.15  0.20 -0.15
MR2  0.10  0.99 -0.07  0.01
MR4  0.18 -0.03  1.00  0.05
MR3 -0.19 -0.05 -0.01  1.00
```

Summary of the 4 factor solution

R code

```
summary(f4)
```

```
Factor analysis with Call: fa(r = prq[8:91], nfactors = 4)

Test of the hypothesis that 4 factors are sufficient.
The degrees of freedom for the model is 3156  and the objective function was  280.09
The number of observations was  75  with Chi Square =  11903.94  with prob <  0

The root mean square of the residuals (RMSA) is  0.09
The df corrected root mean square of the residuals is  0.1

Tucker Lewis Index of factoring reliability =  0
RMSEA index =  0.192  and the 10 % confidence intervals are  0.19 0.197
BIC =  -1722.05
 With factor correlations of
      MR1   MR2   MR4   MR3
MR1  1.00  0.15  0.18 -0.17
MR2  0.15  1.00 -0.03 -0.04
MR4  0.18 -0.03  1.00  0.01
MR3 -0.17 -0.04  0.01  1.00
```

Also try a cluster analysis

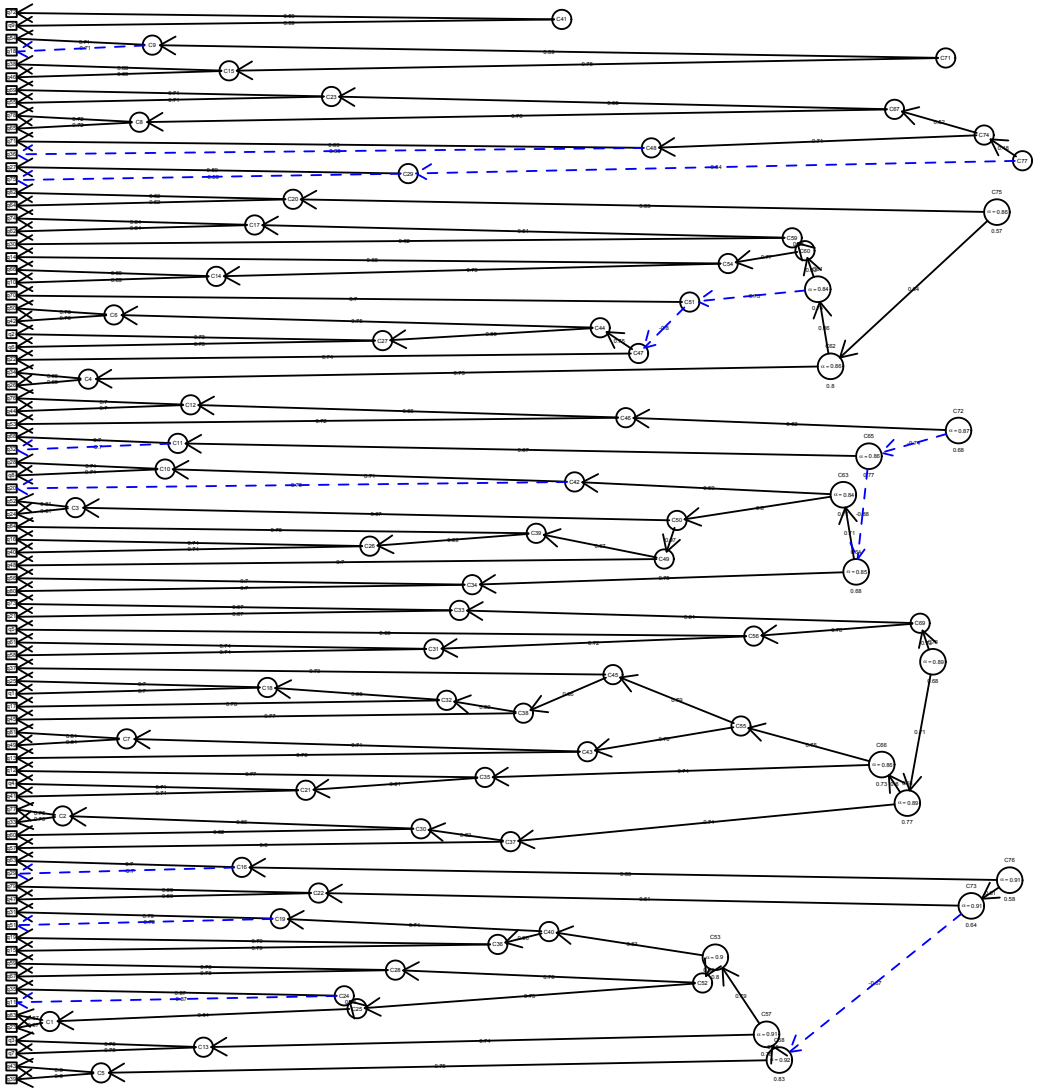
R code

```
ic <-  iclust(prq[8:91])
summary(ic)
```

```
ICLUST (Item Cluster Analysis)Call: iclust(r.mat = prq[8:91])
ICLUST

Purified Alpha:
 C76  C70  C72  C75  C77  C71  C41
0.91 0.89 0.87 0.86 0.72 0.69 0.47

 Guttman Lambda6*
 C76  C70  C72  C75  C77  C71  C41
0.99 0.99 0.98 0.98 0.96 0.96 0.94
Original Beta:
 C76  C70  C72  C75  C77  C71  C41
0.58 0.68 0.68 0.57 0.45 0.58 0.47
Cluster size:
C76 C70 C72 C75 C77 C71 C41
 18  20  15  16   9   4   2

Purified scale intercorrelations
 reliabilities on diagonal
 correlations corrected for attenuation above diagonal:
      C76    C70    C72    C75    C77    C71   C41
C76  0.91 -0.318 -0.467  0.339 -0.271  0.270 -0.29
C70 -0.29  0.891 -0.042  0.013 -0.539  0.130  0.42
C72 -0.42 -0.037  0.875 -0.051  0.418  0.356 -0.37
C75  0.30  0.011 -0.044  0.859  0.134  0.437  0.13
C77 -0.22 -0.431  0.331  0.105  0.716 -0.064 -0.32
C71  0.21  0.102  0.277  0.337 -0.045  0.691 -0.35
C41 -0.19  0.272 -0.240  0.082 -0.184 -0.198  0.47
```

The cluster solution

ICLUST



Compare the solutions

```
> factor.congruence(list(f4,p4,ic))
```

```
factor.congruence(list(f4,p4,ic))
      MR1   MR2   MR4   MR3   RC1   RC2   RC4   RC3   C76   C70   C72   C75   C77   C71
C41
MR1  1.00  0.06  0.12 -0.11  0.99  0.15  0.20 -0.15 -0.93  0.30  0.52 -0.32  0.38 -0.43  0.39
MR2  0.06  1.00 -0.05  0.00  0.10  0.99 -0.07  0.01 -0.23  0.97 -0.09  0.01 -0.71  0.22  0.49
MR4  0.12 -0.05  1.00  0.02  0.18 -0.03  1.00  0.05 -0.28 -0.04  0.90 -0.01  0.40  0.58 -0.53
MR3 -0.11  0.00  0.02  1.00 -0.19 -0.05 -0.01  1.00  0.32 -0.04 -0.05  0.97  0.21  0.48  0.06
RC1  0.99  0.10  0.18 -0.19  1.00  0.19  0.26 -0.23 -0.97  0.34  0.57 -0.39  0.34 -0.41  0.36
RC2  0.15  0.99 -0.03 -0.05  0.19  1.00 -0.05 -0.04 -0.32  0.98 -0.03 -0.05 -0.68  0.16  0.52
RC4  0.20 -0.07  1.00 -0.01  0.26 -0.05  1.00  0.02 -0.35 -0.04  0.93 -0.05  0.45  0.52 -0.51
RC3 -0.15  0.01  0.05  1.00 -0.23 -0.04  0.02  1.00  0.35 -0.04 -0.05  0.98  0.20  0.52  0.03
C76 -0.93 -0.23 -0.28  0.32 -0.97 -0.32 -0.35  0.35  1.00 -0.44 -0.61  0.50 -0.22  0.32 -0.32
C70  0.30  0.97 -0.04 -0.04  0.34  0.98 -0.04 -0.04 -0.44  1.00  0.02 -0.09 -0.57  0.07  0.56
C72  0.52 -0.09  0.90 -0.05  0.57 -0.03  0.93 -0.05 -0.61  0.02  1.00 -0.16  0.52  0.26 -0.30
C75 -0.32  0.01 -0.01  0.97 -0.39 -0.05 -0.05  0.98  0.50 -0.09 -0.16  1.00  0.07  0.56 -0.01
C77  0.38 -0.71  0.40  0.21  0.34 -0.68  0.45  0.20 -0.22 -0.57  0.52  0.07  1.00 -0.02 -0.40
C71 -0.43  0.22  0.58  0.48 -0.41  0.16  0.52  0.52  0.32  0.07  0.26  0.56 -0.02  1.00 -0.33
C41  0.39  0.49 -0.53  0.06  0.36  0.52 -0.51  0.03 -0.32  0.56 -0.30 -0.01 -0.40 -0.33  1.00
>
```

Combine the factor scores with the empirical scores

R code

```
scores.df <- data.frame(f4$scores,prq.scores$scores)
lowerCor(scores.df)
```

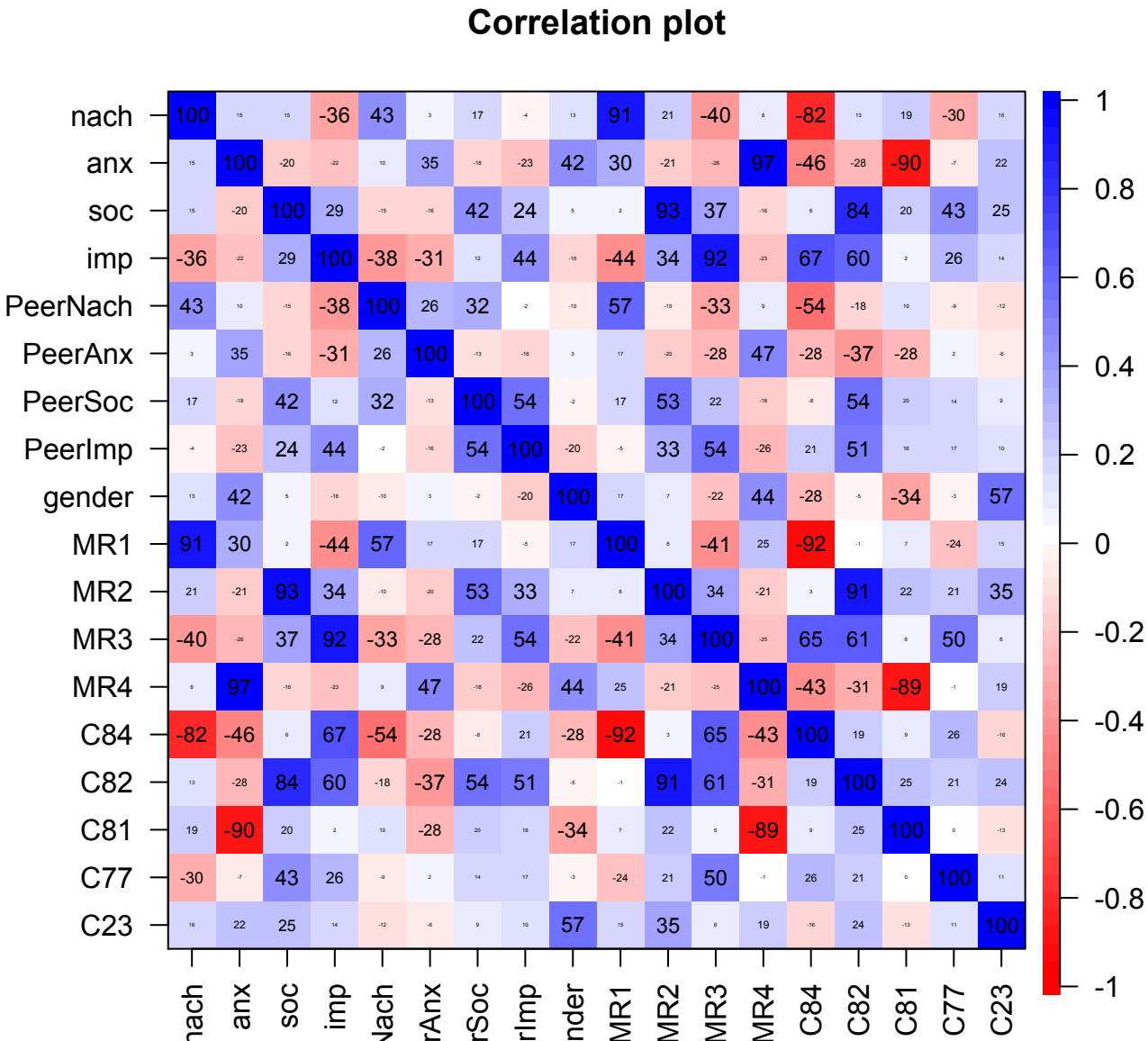
```
             MR1   MR2   MR4   MR3   nach  anxty scblt impls PrNch PrAnx PerSc PrImp gendr
MR1          1.00
MR2          0.16  1.00
MR4          0.20 -0.02  1.00
MR3         -0.20 -0.02  0.02  1.00
nach         0.26  0.94 -0.06 -0.10  1.00
anxiety     -0.25  0.06  0.20  0.92 -0.03  1.00
sociability  0.94  0.21  0.28 -0.27  0.29 -0.26  1.00
impulsivity  0.34 -0.24  0.92 -0.07 -0.22  0.05  0.37  1.00
PeerNach    -0.05  0.19 -0.22  0.06  0.23  0.11  0.00 -0.29  1.00
PeerAnx     -0.25  0.01  0.16  0.54 -0.04  0.59 -0.16  0.05  0.21  1.00
PeerSoc      0.54 -0.19  0.18 -0.12 -0.17 -0.18  0.55  0.35 -0.08 -0.10  1.00
PeerImp      0.22 -0.25  0.42  0.11 -0.26  0.13  0.23  0.50 -0.30 -0.03  0.29  1.00
gender      -0.05 -0.04  0.10  0.13 -0.08  0.18  0.06  0.07  0.00  0.37  0.09  0.05  1.00
```

Compare original, factors and clusters

```
> fkeys <- factor2cluster(f4)
> ckeys <- cluster2keys(ic)
> all.keys <- cbind(prq.keys,fkeys,ckeys)
> all.scores <- scoreItems(all.keys,prq)
> lowerMat(all.scores$cor)
```

```
Coefficients and bootstrapped confidence intervals
          nach  anx   soc   imp   PrNch PrAnx PerSc PrImp gendr MR1   MR2   MR3   MR4   C84
nach       1.00
anx        0.15  1.00
soc        0.15 -0.20  1.00
imp       -0.36 -0.22  0.29  1.00
PeerNach   0.43  0.10 -0.15 -0.38  1.00
PeerAnx    0.03  0.35 -0.16 -0.31  0.26  1.00
PeerSoc    0.17 -0.18  0.42  0.12  0.32 -0.13  1.00
PeerImp   -0.04 -0.23  0.24  0.44 -0.02 -0.16  0.54  1.00
gender     0.13  0.42  0.05 -0.16 -0.10  0.03 -0.02 -0.20  1.00
MR1        0.91  0.30  0.02 -0.44  0.57  0.17  0.17 -0.05  0.17  1.00
MR2        0.21 -0.21  0.93  0.34 -0.10 -0.20  0.53  0.33  0.07  0.08  1.00
MR3       -0.40 -0.26  0.37  0.92 -0.33 -0.28  0.22  0.54 -0.22 -0.41  0.34  1.00
MR4        0.08  0.97 -0.16 -0.23  0.09  0.47 -0.18 -0.26  0.44  0.25 -0.21 -0.25  1.00
C84       -0.82 -0.46  0.06  0.67 -0.54 -0.28 -0.08  0.21 -0.28 -0.92  0.03  0.65 -0.43  1.00
C82        0.13 -0.28  0.84  0.60 -0.18 -0.37  0.54  0.51 -0.05 -0.01  0.91  0.61 -0.31  0.19
C81        0.19 -0.90  0.20  0.02  0.10 -0.28  0.20  0.16 -0.34  0.07  0.22  0.06 -0.89  0.09
C77       -0.30 -0.07  0.43  0.26 -0.09  0.02  0.14  0.17 -0.03 -0.24  0.21  0.50 -0.01  0.26
C23        0.16  0.22  0.25  0.14 -0.12 -0.06  0.09  0.10  0.57  0.15  0.35  0.06  0.19 -0.16
     C82   C81   C77   C23
C82  1.00
C81  0.25  1.00
C77  0.21  0.00  1.00
C23  0.24 -0.13  0.11  1.00
```

The correlations between rational keying, peer ratings factors and clusters



First make a dictionary

A “dictionary" is just a data frame where the row names are the item labels, and the columns are whatever one wants.
Given the length of the items we can `abbreviate` or just select a `substr`ing.
We already have a dictionary and now we want to shorten the text

R code

```
prq.dictionary <-
"https://personality-project.org/courses/405/prq.dictionary.csv"
prq.dictionary <- read.file(prq.dictionary)
abbrev<- substr(prq.dictionary$Content,1,30)
prq.dictionary <- data.frame(prq.dictionary, short=abbrev)
rownames(prq.dictionary) <- prq.dictionary[,1]
```

Factors of PRQ sorted by loadings – F1

```
fa.lookup(f4,prq.dictionary[,c(1,3)])
      MR1   MR2   MR4   MR3   h2  com Item                              short
q35  0.79 -0.01  0.05 -0.12 0.68 1.05  q35  I have a large social network
q11 -0.78 -0.05  0.03  0.08 0.64 1.03  q11  I tend to avoid social situat
q3   0.76  0.10 -0.17  0.16 0.58 1.22  q3   I like to meet new people in
q83  0.76  0.22  0.03 -0.13 0.73 1.22  q83  I am a very sociable person
q39  0.72 -0.04  0.00  0.15 0.49 1.09  q39  Id rather spend time with oth
q23  0.71  0.11  0.08 -0.09 0.59 1.10  q23  I make friends easily
q43  0.61  0.30  0.14  0.08 0.55 1.62  q43  I am happier when Im around o
q51 -0.59 -0.07 -0.15  0.13 0.46 1.25  q51  People are more likely to ini
q67  0.55  0.05  0.26 -0.07 0.44 1.49  q67  I am always willing to attend
q56  0.54 -0.01  0.17  0.22 0.36 1.55  q56  I often and actively express
q59  0.53 -0.21  0.06 -0.08 0.32 1.39  q59  I prefer large crowded partie
q19  0.52  0.19  0.15 -0.14 0.44 1.62  q19  I am good at maintaining a li
q7   0.50  0.09 -0.08 -0.02 0.27 1.12  q7   I can easily start conversati
q79 -0.49  0.36  0.09  0.00 0.30 1.92  q79  When given the choice, I will
q47 -0.46  0.03 -0.14  0.08 0.26 1.26  q47  I enjoy being alone
q71  0.41 -0.29  0.05 -0.10 0.24 1.96  q71  I dont understand how people
q15  0.40 -0.09 -0.06 -0.26 0.25 1.91  q15  I tend to lead the conversati
q63 -0.38 -0.05 -0.04  0.24 0.25 1.74  q63  A good night for me is readin
q9   0.27  0.27 -0.20 -0.03 0.20 2.84  q9   I am a good multi tasker
```

Factors of PRQ sorted by loadings – F2

```
fa.lookup(f4,prq.dictionary[,c(1,3)])
      MR1   MR2   MR4   MR3   h2  com Item                                short
q1   0.20  0.55  0.08 -0.15 0.42 1.48  q1   I love to seek out new challe
q61  0.04  0.54 -0.09 -0.05 0.32 1.09  q61  I experience great joy when m
q49  0.21  0.54  0.04  0.05 0.37 1.33  q49  The joy of success is worth t
q25  0.26  0.54 -0.07 -0.04 0.41 1.50  q25  If I fail, I keep trying unti
q73 -0.14  0.51  0.15 -0.03 0.27 1.34  q73  I set long term and sizeable
q78  0.19 -0.50 -0.03  0.38 0.39 2.20  q78  I tend to back away from task
q45  0.06  0.47  0.02 -0.11 0.25 1.14  q45  I prefer challenging tasks to
q27 -0.29  0.46  0.13  0.01 0.25 1.89  q27  I tend to enjoy small groups
q58 -0.01  0.43  0.18  0.08 0.22 1.42  q58  I prefer to work in relaxed e
q69  0.29 -0.43  0.24  0.15 0.33 2.72  q69  I tend to procrastinate and w
q12 -0.16  0.41 -0.19  0.05 0.23 1.81  q12  I weigh all the options caref
q5  -0.08  0.41  0.13  0.18 0.21 1.72  q5   Personal satisfaction is the
q57  0.24  0.40 -0.12 -0.03 0.26 1.88  q57  I always reach the goals I se
q65  0.09 -0.39  0.10  0.32 0.27 2.21  q65  I tend to have trouble gettin
q37  0.08  0.36  0.20 -0.16 0.22 2.11  q37  I get bored if a task is not
q21 -0.09  0.33 -0.10  0.24 0.18 2.22  q21  I am a perfectionist
q53  0.11 -0.27  0.19  0.19 0.16 3.13  q53  I only work as hard as I have
q75  0.16 -0.25  0.05 -0.05 0.08 1.89  q75  I work better when there are
q29  0.18  0.21  0.10 -0.11 0.12 2.97  q29  I seek the enjoyment of winni
>
```

Factors of PRQ sorted by loadings – F3

```
fa.lookup(f4,prq.dictionary[,c(1,3)])
      MR1   MR2   MR4   MR3   h2  com Item                              short
q24  0.12  0.04  0.71 -0.01 0.54 1.07  q24  I often change my plans at th
q40 -0.02  0.06  0.70 -0.12 0.50 1.08  q40  I act on sudden urges
q52  0.02 -0.13  0.67  0.11 0.49 1.14  q52  I often get sidetracked in th
q38 -0.35  0.10  0.60  0.15 0.45 1.82  q38  I often have unwanted and/or
q8   0.09 -0.16  0.60  0.08 0.42 1.23  q8   I say things that I regret la
q28  0.03 -0.18  0.56  0.08 0.36 1.26  q28  I dislike planning ahead
q84  0.21  0.03  0.55 -0.08 0.41 1.34  q84  I am an impulsive person
q44  0.21 -0.16  0.54  0.27 0.45 2.03  q44  I often regret decisions beca
q32  0.15  0.14  0.50  0.11 0.33 1.45  q32  I indulge in my desires on a
q68 -0.09  0.24 -0.49  0.04 0.32 1.56  q68  I always think before I act
q76  0.20  0.07  0.48  0.08 0.32 1.42  q76  I sometimes look back and don
q72  0.24  0.08 -0.48  0.24 0.29 2.08  q72  I always stick to plans
q48  0.05  0.24  0.48 -0.19 0.33 1.88  q48  I tend to act on my gut feeli
q16  0.34  0.11  0.45 -0.23 0.47 2.59  q16  I tend to make decisions quic
q20 -0.06  0.35 -0.41  0.17 0.33 2.36  q20  I plan my activities in advan
q46 -0.30  0.16  0.38 -0.02 0.20 2.29  q46  I often have difficulty sleep
q80  0.30 -0.22  0.35 -0.07 0.29 2.77  q80  I often say the first thing t
q54 -0.24  0.21  0.31  0.21 0.21 3.60  q54  I feel tension in my body or
q36 -0.09  0.20 -0.22 -0.02 0.10 2.39  q36  When working on a necessary t
>
```

Factors of PRQ sorted by loadings – F4

```
fa.lookup(f4,prq.dictionary[,c(1,3)])
      MR1   MR2   MR4   MR3   h2  com Item                                short
q6  -0.13 -0.10  0.04  0.67 0.51 1.13  q6  I dont handle stress well
q50 -0.01  0.02  0.01  0.63 0.40 1.00  q50  Even in non stressful situati
q42 -0.16  0.02  0.06  0.63 0.45 1.15  q42  Even trivial problems greatly
q66  0.20 -0.07 -0.06  0.62 0.39 1.25  q66  I worry about what others thi
q2  -0.17 -0.15  0.00  0.60 0.47 1.30  q2  I get nervous very easily
q10  0.02 -0.03  0.05  0.57 0.33 1.02  q10 I am easily bothered by negati
q62  0.02  0.19 -0.04  0.56 0.34 1.24  q62  A small unpleasant event can
q22 -0.03  0.26 -0.16  0.55 0.39 1.63  q22 I feel stressed when I have a
q34 -0.11  0.26 -0.01  0.52 0.35 1.59  q34 I have a hard time forgetting
q26 -0.14  0.27  0.26  0.50 0.40 2.31  q26 I often feel anxious about fut
q64  0.12 -0.02 -0.42  0.45 0.36 2.13  q64  I dislike changing establishe
q31  0.33  0.11  0.26 -0.45 0.48 2.63  q31  I tend to talk a lot in large
q82  0.30 -0.09 -0.08  0.45 0.24 1.94  q82 I am more emotional than my fr
q30 -0.13  0.02  0.25  0.44 0.28 1.79  q30 I often feel tense, nauseous,
q70  0.23  0.28  0.13 -0.41 0.39 2.66  q70  I bounce back quickly from un
q18  0.25 -0.29 -0.28 -0.41 0.37 3.38  q18 I rarely feel tense
q74  0.16  0.27  0.14  0.41 0.27 2.37  q74  I tend to dwell on obstacles
q55  0.35 -0.16  0.28  0.36 0.34 3.29  q55  Ill spend time talking to a f
q14 -0.20 -0.04  0.12  0.27 0.14 2.33  q14 Measures of skill or intellige
>
```

Show the items for the clusters

R code

```
fa.lookup(ic,prq.dictionary[,c(1,3)])
```

	C76	C70	C72	C75	C77	C71	C41	Item	short
q83	-0.83	0.37	0.30	-0.27	0.04	-0.20	0.18	q83	I am a very sociable person
q35	-0.81	0.19	0.36	-0.27	0.29	-0.21	0.15	q35	I have a large social network
q11	0.77	-0.24	-0.30	0.25	-0.21	0.25	-0.21	q11	I tend to avoid social situat
q51	0.74	-0.20	-0.34	0.23	-0.14	0.10	-0.02	q51	People are more likely to ini
q23	-0.71	0.28	0.36	-0.23	0.16	-0.26	0.12	q23	I make friends easily
q67	-0.68	0.20	0.42	-0.17	0.23	0.02	0.04	q67	I am always willing to attend
q39	-0.67	0.14	0.27	0.01	0.28	-0.16	0.35	q39	Id rather spend time with oth
q3	-0.67	0.27	0.15	-0.02	0.17	-0.24	0.22	q3	I like to meet new people in
q43	-0.67	0.46	0.33	-0.08	0.18	-0.08	0.22	q43	I am happier when Im around o
q19	-0.65	0.28	0.36	-0.19	-0.09	-0.12	0.16	q19	I am good at maintaining a li
q31	-0.61	0.16	0.36	-0.44	-0.07	-0.07	-0.06	q31	I tend to talk a lot in large
q59	-0.60	-0.06	0.29	-0.13	0.26	-0.10	0.12	q59	I prefer large crowded partie
q47	0.58	-0.09	-0.29	0.15	-0.11	0.05	-0.11	q47	I enjoy being alone
q63	0.56	-0.16	-0.15	0.30	-0.14	0.21	-0.10	q63	A good night for me is readin
q7	-0.53	0.21	0.14	-0.14	0.09	-0.02	0.24	q7	I can easily start conversati
q15	-0.49	0.01	0.11	-0.31	0.04	-0.12	-0.08	q15	I tend to lead the conversati
q79	0.45	0.20	-0.15	0.09	-0.23	0.30	-0.07	q79	When given the choice, I will
q29	-0.33	0.31	0.12	-0.17	0.12	-0.03	0.07	q29	I seek the enjoyment of winni
q81	-0.29	0.75	0.11	-0.01	-0.31	0.04	0.10	q81	I believe that if something i
q17	-0.13	0.71	-0.09	0.14	-0.39	0.04	0.17	q17	I have high standards for the
q33	-0.28	0.70	0.14	0.03	-0.26	0.15	0.39	q33	I find myself needing to achi
q25	-0.32	0.64	0.03	-0.11	-0.21	0.02	0.26	q25	If I fail, I keep trying unti
q4	-0.12	0.63	-0.38	-0.13	-0.33	0.00	0.20	q4	I am thoughtful and deliberat
q13	-0.20	0.62	-0.10	0.01	-0.34	0.05	0.26	q13	I like to go the extra mile o
q41	-0.20	0.61	-0.07	-0.06	-0.27	0.12	0.33	q41	I always make sure anything a
q77	-0.18	0.61	-0.06	0.12	-0.29	0.15	0.39	q77	I always see projects through
q1	-0.39	0.61	0.10	-0.14	-0.23	0.07	0.22	q1	I love to seek out new challe
q60	-0.18	0.61	-0.07	0.06	-0.37	0.05	0.27	q60	I stay on task until a projec
q49	-0.29	0.60	0.09	0.03	-0.20	-0.08	0.02	q49	The joy of success is worth t
q61	-0.17	0.60	-0.12	-0.07	-0.28	-0.07	0.08	q61	I experience great joy when m

Cluster 2

	C76	C70	C72	C75	C77	C71	C41	Item	short
q81	-0.29	0.75	0.11	-0.01	-0.31	0.04	0.10	q81	I believe that if something i
q17	-0.13	0.71	-0.09	0.14	-0.39	0.04	0.17	q17	I have high standards for the
q33	-0.28	0.70	0.14	0.03	-0.26	0.15	0.39	q33	I find myself needing to achi
q25	-0.32	0.64	0.03	-0.11	-0.21	0.02	0.26	q25	If I fail, I keep trying unti
q4	-0.12	0.63	-0.38	-0.13	-0.33	0.00	0.20	q4	I am thoughtful and deliberat
q13	-0.20	0.62	-0.10	0.01	-0.34	0.05	0.26	q13	I like to go the extra mile o
q41	-0.20	0.61	-0.07	-0.06	-0.27	0.12	0.33	q41	I always make sure anything a
q77	-0.18	0.61	-0.06	0.12	-0.29	0.15	0.39	q77	I always see projects through
q1	-0.39	0.61	0.10	-0.14	-0.23	0.07	0.22	q1	I love to seek out new challe
q60	-0.18	0.61	-0.07	0.06	-0.37	0.05	0.27	q60	I stay on task until a projec
q49	-0.29	0.60	0.09	0.03	-0.20	-0.08	0.02	q49	The joy of success is worth t
q61	-0.17	0.60	-0.12	-0.07	-0.28	-0.07	0.08	q61	I experience great joy when m
q73	-0.06	0.54	0.01	-0.01	-0.22	0.29	-0.02	q73	I set long term and sizeable
q45	-0.12	0.52	0.07	-0.08	-0.25	-0.02	0.31	q45	I prefer challenging tasks to
q57	-0.28	0.51	-0.01	-0.06	-0.22	-0.05	0.30	q57	I always reach the goals I se
q12	0.16	0.46	-0.34	0.02	-0.09	0.01	0.02	q12	I weigh all the options caref
q58	-0.11	0.44	0.10	0.07	-0.13	0.14	-0.03	q58	I prefer to work in relaxed e
q37	-0.21	0.42	0.22	-0.16	-0.19	0.10	-0.05	q37	I get bored if a task is not
q5	0.02	0.41	0.09	0.19	-0.23	0.06	-0.07	q5	Personal satisfaction is the
q21	0.06	0.36	-0.15	0.27	-0.21	0.11	0.06	q21	I am a perfectionist

Cluster 3

	C76	C70	C72	C75	C77	C71	C41	Item		short
q40	-0.18	0.05	0.72	-0.08	0.06	0.25	-0.30	q40	I act on sudden urges	
q24	-0.28	0.08	0.71	-0.06	0.31	0.26	-0.19	q24	I often change my plans at th	
q8	-0.18	-0.12	0.68	0.05	0.25	0.24	-0.31	q8	I say things that I regret la	
q84	-0.33	0.05	0.67	-0.13	0.13	0.08	-0.11	q84	I am an impulsive person	
q28	-0.02	-0.14	0.64	0.06	0.37	0.21	-0.24	q28	I dislike planning ahead	
q32	-0.23	0.14	0.63	0.07	0.02	0.30	-0.02	q32	I indulge in my desires on a	
q52	-0.12	-0.17	0.62	0.08	0.38	0.33	-0.32	q52	I often get sidetracked in th	
q44	-0.26	-0.11	0.61	0.21	0.38	0.29	-0.31	q44	I often regret decisions beca	
q16	-0.53	0.17	0.59	-0.24	0.11	0.06	0.04	q16	I tend to make decisions quic	
q76	-0.31	0.08	0.57	0.04	0.20	0.24	-0.12	q76	I sometimes look back and don	
q80	-0.35	-0.16	0.56	-0.11	0.18	-0.06	0.03	q80	I often say the first thing t	
q68	0.20	0.28	-0.55	0.02	-0.13	-0.19	0.16	q68	I always think before I act	
q56	-0.46	0.12	0.48	0.10	0.17	-0.11	0.12	q56	I often and actively express	
q20	0.08	0.29	-0.47	0.20	-0.38	0.01	0.27	q20	I plan my activities in advan	
q48	-0.27	0.23	0.47	-0.16	-0.02	0.29	-0.09	q48	I tend to act on my gut feeli	
q42	0.28	-0.07	-0.06	0.70	0.01	0.31	-0.08	q42	Even trivial problems greatly	
q6	0.34	-0.15	-0.03	0.69	0.16	0.28	-0.03	q6	I dont handle stress well	
q50	0.15	0.00	0.00	0.68	0.04	0.36	0.11	q50	Even in non stressful situati	
q2	0.35	-0.19	-0.10	0.65	0.18	0.32	-0.09	q2	I get nervous very easily	
q10	0.17	-0.03	0.04	0.62	0.11	0.17	0.10	q10	I am easily bothered by negati	
q66	0.03	-0.01	-0.02	0.60	0.34	0.06	0.15	q66	I worry about what others thi	
q62	0.13	0.17	-0.06	0.59	-0.04	0.08	0.02	q62	A small unpleasant event can	
q34	0.18	0.22	-0.09	0.59	-0.11	0.23	0.18	q34	I have a hard time forgetting	
q22	0.19	0.22	-0.20	0.58	-0.08	0.26	0.12	q22	I feel stressed when I have a	
q70	-0.40	0.35	0.20	-0.56	-0.01	-0.10	0.01	q70	I bounce back quickly from un	
q26	0.16	0.20	0.17	0.55	-0.01	0.36	0.07	q26	I often feel anxious about fut	
q30	0.20	-0.05	0.20	0.53	0.07	0.34	0.04	q30	I often feel tense, nauseous,	
q74	-0.12	0.25	0.12	0.44	0.08	0.22	-0.10	q74	I tend to dwell on obstacles	
q82	-0.09	-0.04	0.13	0.44	0.09	-0.03	0.17	q82	I am more emotional than my fr	
q64	0.18	0.03	-0.29	0.43	0.02	-0.17	0.26	q64	I dislike changing establishe	
q14	0.21	-0.10	-0.01	0.35	0.10	0.21	-0.14	q14	Measures of skill or intellige	
q69	-0.17	-0.29	0.36	0.01	0.72	-0.09	-0.22	q69	I tend to procrastinate and w	

Cluster 4

	C76	C70	C72	C75	C77	C71	C41	Item	short
q42	0.28	-0.07	-0.06	0.70	0.01	0.31	-0.08	q42	Even trivial problems greatly
q6	0.34	-0.15	-0.03	0.69	0.16	0.28	-0.03	q6	I dont handle stress well
q50	0.15	0.00	0.00	0.68	0.04	0.36	0.11	q50	Even in non stressful situati
q2	0.35	-0.19	-0.10	0.65	0.18	0.32	-0.09	q2	I get nervous very easily
q10	0.17	-0.03	0.04	0.62	0.11	0.17	0.10	q10	I am easily bothered by negati
q66	0.03	-0.01	-0.02	0.60	0.34	0.06	0.15	q66	I worry about what others thi
q62	0.13	0.17	-0.06	0.59	-0.04	0.08	0.02	q62	A small unpleasant event can
q34	0.18	0.22	-0.09	0.59	-0.11	0.23	0.18	q34	I have a hard time forgetting
q22	0.19	0.22	-0.20	0.58	-0.08	0.26	0.12	q22	I feel stressed when I have a
q70	-0.40	0.35	0.20	-0.56	-0.01	-0.10	0.01	q70	I bounce back quickly from un
q26	0.16	0.20	0.17	0.55	-0.01	0.36	0.07	q26	I often feel anxious about fut
q30	0.20	-0.05	0.20	0.53	0.07	0.34	0.04	q30	I often feel tense, nauseous,
q74	-0.12	0.25	0.12	0.44	0.08	0.22	-0.10	q74	I tend to dwell on obstacles
q82	-0.09	-0.04	0.13	0.44	0.09	-0.03	0.17	q82	I am more emotional than my fr
q64	0.18	0.03	-0.29	0.43	0.02	-0.17	0.26	q64	I dislike changing establishe
q14	0.21	-0.10	-0.01	0.35	0.10	0.21	-0.14	q14	Measures of skill or intellige
q69	-0.17	-0.29	0.36	0.01	0.72	-0.09	-0.22	q69	I tend to procrastinate and w
q65	0.11	-0.30	0.16	0.24	0.67	0.06	-0.08	q65	I tend to have trouble gettin
q78	0.04	-0.46	0.05	0.25	0.62	0.07	-0.05	q78	I tend to back away from task
q36	0.08	0.15	-0.19	0.04	-0.57	-0.12	0.14	q36	When working on a necessary t
q55	-0.22	-0.06	0.38	0.22	0.55	0.01	-0.23	q55	Ill spend time talking to a f
q53	-0.04	-0.21	0.22	0.10	0.50	0.01	-0.18	q53	I only work as hard as I have
q71	-0.38	-0.17	0.18	-0.13	0.44	-0.22	0.01	q71	I dont understand how people
q27	0.19	0.34	-0.02	0.03	-0.42	0.23	0.07	q27	I tend to enjoy small groups
q75	-0.17	-0.18	0.10	-0.09	0.40	0.05	0.00	q75	I work better when there are
q38	0.17	-0.02	0.40	0.24	0.03	0.75	-0.13	q38	I often have unwanted and/or
q18	-0.19	-0.17	-0.08	-0.45	0.04	-0.74	0.15	q18	I rarely feel tense
q54	0.12	0.10	0.09	0.21	0.02	0.71	-0.16	q54	I feel tension in my body or
q46	0.14	0.05	0.24	0.07	-0.14	0.62	-0.15	q46	I often have difficulty sleep
q72	-0.05	0.14	-0.34	0.18	-0.07	-0.18	0.79	q72	I always stick to plans
q9	-0.26	0.30	-0.05	-0.05	-0.23	-0.15	0.77	q9	I am a good multi tasker

Cluster 5

	C76	C70	C72	C75	C77	C71	C41	Item		short
q69	-0.17	-0.29	0.36	0.01	0.72	-0.09	-0.22	q69	I tend to procrastinate and w	
q65	0.11	-0.30	0.16	0.24	0.67	0.06	-0.08	q65	I tend to have trouble gettin	
q78	0.04	-0.46	0.05	0.25	0.62	0.07	-0.05	q78	I tend to back away from task	
q36	0.08	0.15	-0.19	0.04	-0.57	-0.12	0.14	q36	When working on a necessary t	
q55	-0.22	-0.06	0.38	0.22	0.55	0.01	-0.23	q55	Ill spend time talking to a f	
q53	-0.04	-0.21	0.22	0.10	0.50	0.01	-0.18	q53	I only work as hard as I have	
q71	-0.38	-0.17	0.18	-0.13	0.44	-0.22	0.01	q71	I dont understand how people	
q27	0.19	0.34	-0.02	0.03	-0.42	0.23	0.07	q27	I tend to enjoy small groups	
q75	-0.17	-0.18	0.10	-0.09	0.40	0.05	0.00	q75	I work better when there are	
q38	0.17	-0.02	0.40	0.24	0.03	0.75	-0.13	q38	I often have unwanted and/or	
q18	-0.19	-0.17	-0.08	-0.45	0.04	-0.74	0.15	q18	I rarely feel tense	
q54	0.12	0.10	0.09	0.21	0.02	0.71	-0.16	q54	I feel tension in my body or	
q46	0.14	0.05	0.24	0.07	-0.14	0.62	-0.15	q46	I often have difficulty sleep	
q72	-0.05	0.14	-0.34	0.18	-0.07	-0.18	0.79	q72	I always stick to plans	
q9	-0.26	0.30	-0.05	-0.05	-0.23	-0.15	0.77	q9	I am a good multi tasker	

Empirical scale construction

1. Identify those items that most correlate with the criteria
   - Form item composites based upon those items
2. `best.scales` will do this
   - bs <- bestScales(prq[3:91],colnames(prq[3:7]), dictionary=prq.dictionary[3],n.item=20)

Empirical 1

```
Call = bestScales(x = prq[3:91], criteria = colnames(prq[3:7]), n.item = 20,
    dictionary = prq.dictionary[2])
The items most correlated with the criteria yield r's of
            correlation n.items
NeedAch            0.65      20
Anxiety            0.68      20
Sociability        0.69      20
Impulsivity        0.66      20
Gender             0.58      20


The best items, their correlations and content  are
$NeedAch
     NeedAch                              short
q60     0.36  I stay on task until a projec
q68     0.33  I always think before I act
q13     0.32  I like to go the extra mile o
q32    -0.31  I indulge in my desires on a
q69    -0.30  I tend to procrastinate and w
q65    -0.29  I tend to have trouble gettin
q6      0.29 I dont handle stress well
q80    -0.28  I often say the first thing t
q22     0.26 I feel stressed when I have a
q53    -0.25  I only work as hard as I have
q75     0.24  I work better when there are
q84    -0.23  I am an impulsive person
q10     0.22 I am easily bothered by negati
q73     0.21  I set long term and sizeable
q19     0.21  I am good at maintaining a li
q76    -0.21  I sometimes look back and don
q24    -0.20  I often change my plans at th
q42     0.20  Even trivial problems greatly
q52    -0.20  I often get sidetracked in th
```

Empirical 2

```
$Anxiety
 Anxiety                              short
q42    0.54  Even trivial problems greatly
q6     0.51 I dont handle stress well
q18   -0.47 I rarely feel tense
q62    0.46  A small unpleasant event can
q63    0.35  A good night for me is readin
q2     0.35 I get nervous very easily
q50    0.32  Even in non stressful situati
q54    0.31  I feel tension in my body or
q21    0.31  I am a perfectionist
q44    0.30  I often regret decisions beca
q30    0.29 I often feel tense, nauseous,
q5     0.29  Personal satisfaction is the
q34    0.29 I have a hard time forgetting
q38    0.28  I often have unwanted and/or
q74    0.28  I tend to dwell on obstacles
q35   -0.28  I have a large social network
q10    0.28 I am easily bothered by negati
q71   -0.28  I dont understand how people
q22    0.27 I feel stressed when I have a
q14    0.27 Measures of skill or intellige
```

Empirical 3

```
$Sociability
  Sociability                          short
q35        0.51  I have a large social network
q39        0.46  Id rather spend time with oth
q3         0.45  I like to meet new people in
q7         0.44  I can easily start conversati
q51       -0.44  People are more likely to ini
q83        0.42  I am a very sociable person
q11       -0.41  I tend to avoid social situat
q73       -0.40  I set long term and sizeable
q31        0.38  I tend to talk a lot in large
q19        0.36  I am good at maintaining a li
q43        0.35  I am happier when Im around o
q36       -0.35  When working on a necessary t
q71        0.33  I dont understand how people
q15        0.33  I tend to lead the conversati
q68       -0.33  I always think before I act
q23        0.32  I make friends easily
q59        0.32  I prefer large crowded partie
q56        0.31  I often and actively express
q60       -0.31  I stay on task until a projec
q79       -0.28  When given the choice, I will
```

Empirical 4

```
$Impulsivity
Impulsivity                              short
q84          0.47  I am an impulsive person
q4          -0.46  I am thoughtful and deliberat
q69          0.45  I tend to procrastinate and w
q32          0.41  I indulge in my desires on a
q52          0.37  I often get sidetracked in th
q40          0.35  I act on sudden urges
q12         -0.33  I weigh all the options caref
q16          0.33  I tend to make decisions quic
q20         -0.32  I plan my activities in advan
q68         -0.30  I always think before I act
q24          0.29  I often change my plans at th
q23          0.28  I make friends easily
q36         -0.28  When working on a necessary t
q77         -0.28  I always see projects through
q57         -0.26  I always reach the goals I se
q56          0.26  I often and actively express
q60         -0.25  I stay on task until a projec
q67          0.25  I am always willing to attend
q76          0.24  I sometimes look back and don
q78          0.24  I tend to back away from task
```

Items predicting gender

R code

```
    Gender                                short
q57  -0.30  I always reach the goals I se
q27   0.30  I tend to enjoy small groups
q5    0.25  Personal satisfaction is the
q77  -0.23  I always see projects through
q54   0.23  I feel tension in my body or
q6    0.23 I dont handle stress well
q55   0.21  Ill spend time talking to a f
q42   0.21  Even trivial problems greatly
q72  -0.21  I always stick to plans
q71  -0.20  I dont understand how people
q45  -0.19  I prefer challenging tasks to
q58   0.19  I prefer to work in relaxed e
q33  -0.19  I find myself needing to achi
q21   0.17  I am a perfectionist
q34   0.17 I have a hard time forgetting
q52   0.16  I often get sidetracked in th
q17  -0.15  I have high standards for the
q75   0.15  I work better when there are
q41  -0.15  I always make sure anything a
q56   0.15  I often and actively express
```

Multiple ways to construct scales

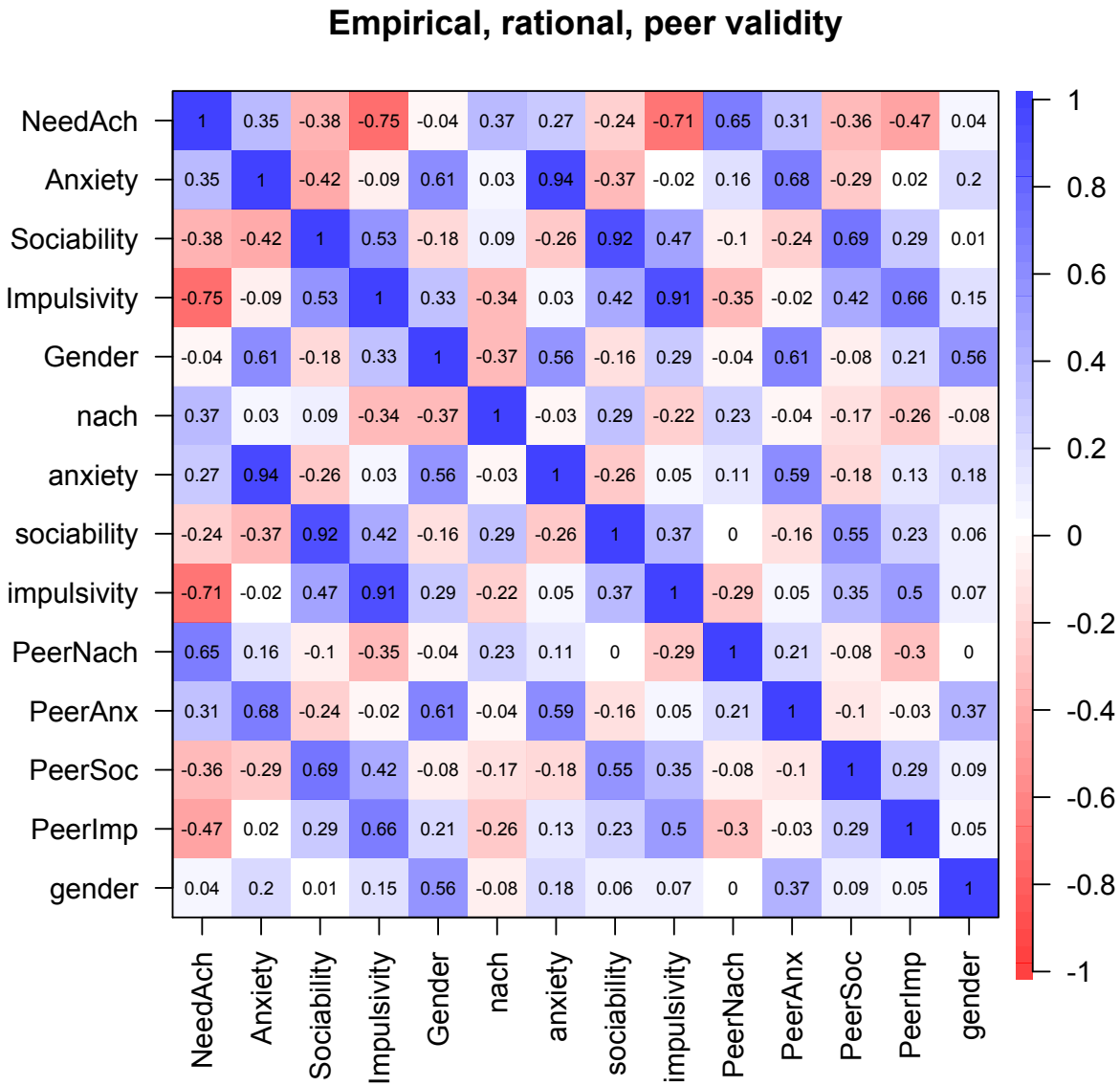
1. Rational/Theoretical
   - Learn Theory
   - Write good items
2. Homogeneous keying
   - Write good items
   - Factor/Cluster analyze
3. Empirical Keys
   - Write good items
   - Select those items that correlate with the criteria

Reliability of various ways of scoring

```
prq.emp <- keys2list(bs$key)
  mixed.key <-  c(prq.emp,prq.keys)
  mixed <- scoreItems(mixed.key,prq)
> mixed
```

```
> mixed
Call: scoreItems(keys = mixed.key, items = prq)
Call: scoreItems(keys = mixed.key, items = prq)

(Unstandardized) Alpha:
      NeedAch Anxiety Sociability Impulsivity Gender nach anxiety sociability impulsivity PeerNach
alpha    0.71    0.84        0.87        0.83    0.6 0.87    0.85        0.87        0.87
1       1       1       1       1

Standard errors of unstandardized Alpha:
      NeedAch Anxiety Sociability Impulsivity Gender nach anxiety sociability impulsivity PeerNach
ASE     0.056   0.035       0.029       0.037  0.073 0.03   0.032       0.03        0.03
NaN     NaN     NaN     NaN     NaN

Average item correlation:
          NeedAch Anxiety Sociability Impulsivity Gender nach anxiety sociability impulsivity PeerI
average.r    0.11     0.2        0.26        0.19   0.07 0.24    0.21        0.24        0.23
NaN     NaN     NaN     NaN     NaN

Median item correlation:
    NeedAch      Anxiety Sociability Impulsivity       Gender         nach      anxiety sociability im
PeerNach      PeerAnx      PeerSoc
      0.110        0.205        0.260        0.209        0.074        0.249        0.227
0.283        0.251           NA           NA           NA
    PeerImp       gender
         NA           NA
```

Show the MMTM matrix graphically – corPlot(mixed$cor,main="Empirical, rational, peer validity",xlas=3))



10 steps: Reprise

1. Specify your theory of relevant constructs
2. Define the population of interest
3. Give items to engaged subjects
4. Enter the data (carefully)
5. Descriptives to double check data entry and subject engagement
6. Find the variance/covariance matrix
7. Reduce its dimensionality through FA, PC, or clustering
8. Score composites (classical or IRT based)
9. Discriminant validity versus other constructs
10. Convergent validity with similar constructs and different methods

Methods of scale construction

1. Empirical
   - MMPI
   - Strong Vocational Interest Blank
2. Rational
   - California Psychological Inventory
3. Theoretical
   - Measures of Need Achievement (e.g., Jackson PI)
4. Homegeneous keying
   - Eysenck Personality Inventory
   - NEO
   - BFI
   - TIPI

Empirical

1. Ask items that discriminate known groups
   - People in general versus specific group
   - Choose items that are maximally independent and that have highest validities
2. Example:
   - MMPI
   - Strong-Campbell
   - sex and ethnic differences in personality and music
3. Problem:
   - What is the meaning of the scale?
   - Need to develop new scale for every new group

Sex differences at item level

Item	effect size
Get overwhelmed by emotions.	0.59
Sympathize with others' feelings.	0.45
Worry about things.	0.43
Feel others' emotions.	0.39
Get stressed out easily.	0.51
Have a soft heart.	0.38
Panic easily	0.50
Inquire about others' well-being.	0.41
Get upset by unpleasant thoughts that come into my mind.	0.38
Get upset easily.	0.37
Am indifferent to the feelings of others.	-0.33
Am not interested in other people's problems.	-0.33
Feel little concern for others.	-0.35
Am not easily bothered by things	-0.35
Love to help others.	0.34
Am not really interested in others.	-0.32
Think of others first.	0.30
Take offense easily.	0.29
Take time out for others.	0.33

Sex differences and music preference

effect size	Item
0.9	Broadway Musicals (e.g. Rent, Cats, Phantom of the Opera)
0.68	Top 40/Pop Vocal Music (e.g. Kelly Clarkson, Madonna, The Black Eyed Peas)
0.65	Broadway, Movie and TV Soundtrack Music in General
0.59	Contemporary Rhythm and Blues (e. g. Whitney Houston, Usher, Alicia Keys)
0.59	Modern Country Music (e.g. Garth Brooks, Dixie Chicks, Tim McGraw)
0.37	Country Music in General
0.37	Movie Soundtracks (e.g. Starwars, Good Will Hunting, Garden State)
0.36	Top 40 Music/Pop in General
0.32	Pop Rock (e.g. Maroon 5, Counting Crows, John Mayer)
0.31	Modern Religious Music (e.g. 4Him, Casting Crowns)
0.3	Soul Rock (e.g. Stevie Wonder, Earth Wind and Fire)
-0.3	Acid Rock (e.g. Pink Floyd, The Doors, Jefferson Airplane)
-0.4	Heavy Metal (e.g. Metallica, Marilyn Manson, System of a Down)

Ethnic differences and music preference

effect size	Item
1.26	Acid Rock (e.g. Pink Floyd, The Doors, Jefferson Airplane)
1	Alternative (e.g. Pearl Jam, Incubus, Radiohead)
0.97	Electronic Music in General
0.91	Rock Music In General
0.87	Jam Bands (e.g. The Grateful Dead, Phish, String Cheese Incident)
0.87	Classic Rock (e.g. The Beatles, The Rolling Stones, Led Zeppelin)
0.85	Country Rock (e.g. The Allman Brothers, Lynyrd Skynyrd)
0.61	Electronic Dance Music (e.g. DJ Tiesto, Paul Van Dyk, Keoki)
0.59	Folk Music in General (e.g. Bob Dylan, Iron and Wine, Simon and Garfunkel)
0.57	Pop Rock (e.g. Maroon 5, Counting Crows, John Mayer)
0.56	Country Music in General
0.51	Bluegrass (e.g. Alison Krauss, Lester Flatt, Nickel Creek)
-0.56	Contemporary Rhythm and Blues (e. g. Whitney Houston, Usher, Alicia Keys)
-0.6	Blues in General (e.g. Ray Charles, Stevie Ray Vaughn, B.B. King)
-0.63	Instrumental Hip-Hop (e.g. DJ Hi-Tek, RJD2, Prefuse 73)
-0.64	Gospel Soul (e.g. Aretha Franklin, Solomon Burke)
-0.67	Soul in General (e.g. Otis Redding, Marvin Gaye)
-0.84	Religious Music in General
-1.04	Soul Rock (e.g. Stevie Wonder, Earth Wind and Fire)
-1.11	Rhythm and Blues in General
-1.43	Religious Gospel (e.g. Andre Crouch, Gospel Quartet)

Rational Keying

1. Ask items with direct content relevance
2. Example: California Psychological Inventory
3. Problems
   - Not all items predict in obvious way
   - Need evidence for validity
   - Easy to fake

Theoretical Keying

1. Ask items with theoretical relevance
2. Example: Jackson Personality Research Form
3. Problems:
   - Theoretical circularity
   - Need evidence for validity

Homogeneous Keying

1. Select items to represent single domain
2. Exclude items based upon internal consistency
3. Examples:
   - 16PF
   - EPI/EPQ,
   - NEO/NEO-PIR
4. Problems
   - Garbage In, Garbage Out
   - Need evidence for validity

Methods of Homogeneous keying

1. Cluster analysis (e.g. `iclust`)
2. Principal Components analysis (e.g., `pca`)
3. Factor analysis (e.g., `fa`)

The Hase and Goldberg and Goldberg studies

1. Hase and Goldberg: a direct comparison of different techniques
   - Differential validity of scale construction
   - Factor analytic
   - Empirical Group discrimination
   - Intuitive theoretical
   - Intuitive rational
   - Stylistic-psychometric
   - Random
2. 200 University Freshman women
3. CPI items and 13 criteria

But compare to Revelle, Dworak & Condon (2021)

Hase and Goldberg: 13 criteria

1. Sorority Membership
2. An experimental measure of conformity
3. Peer ratings of
   - Dominance
   - Sociability
   - Responsibility
   - Psychological Mindedness
   - Femininity
4. Peer ratings of how well known the person is
5. Average number of dates per month
6. College Grade Point Average
7. College Achievement relative to ability
8. College Major
9. College Droput

Does it make a difference?

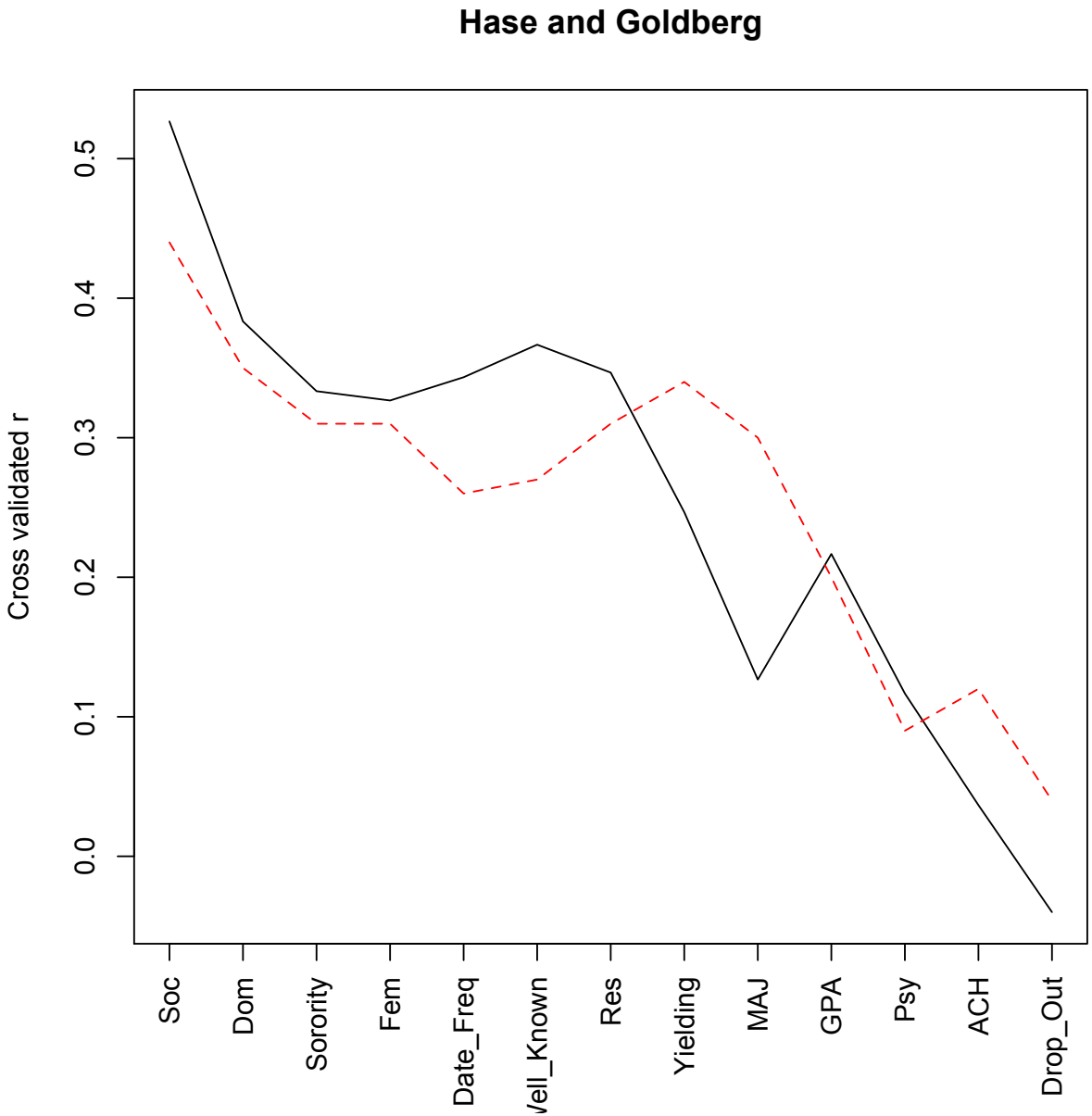
1. Hase and Goldberg (Hase & Goldberg, 1967) No
2. Goldberg (1972) YES

Hase and Goldberg; mean values

Original Hase and Goldberg showed no difference between methods, except that stylistic and random were much worse.

	var	n	mean	sd	median	trimmed	mad	min	max	range	se
Factor	1	13	0.25	0.18	0.27	0.25	0.13	-0.05	0.57	0.62	0.05
Theoretical	2	13	0.25	0.16	0.26	0.25	0.18	0.01	0.52	0.51	0.04
Rational	3	13	0.26	0.16	0.32	0.27	0.09	-0.08	0.49	0.57	0.04
Empirical	4	13	0.26	0.11	0.30	0.26	0.06	0.04	0.44	0.40	0.03
Stylistic	5	13	0.13	0.12	0.11	0.13	0.12	-0.07	0.35	0.42	0.03
Random	6	13	0.10	0.12	0.11	0.10	0.13	-0.08	0.30	0.38	0.03

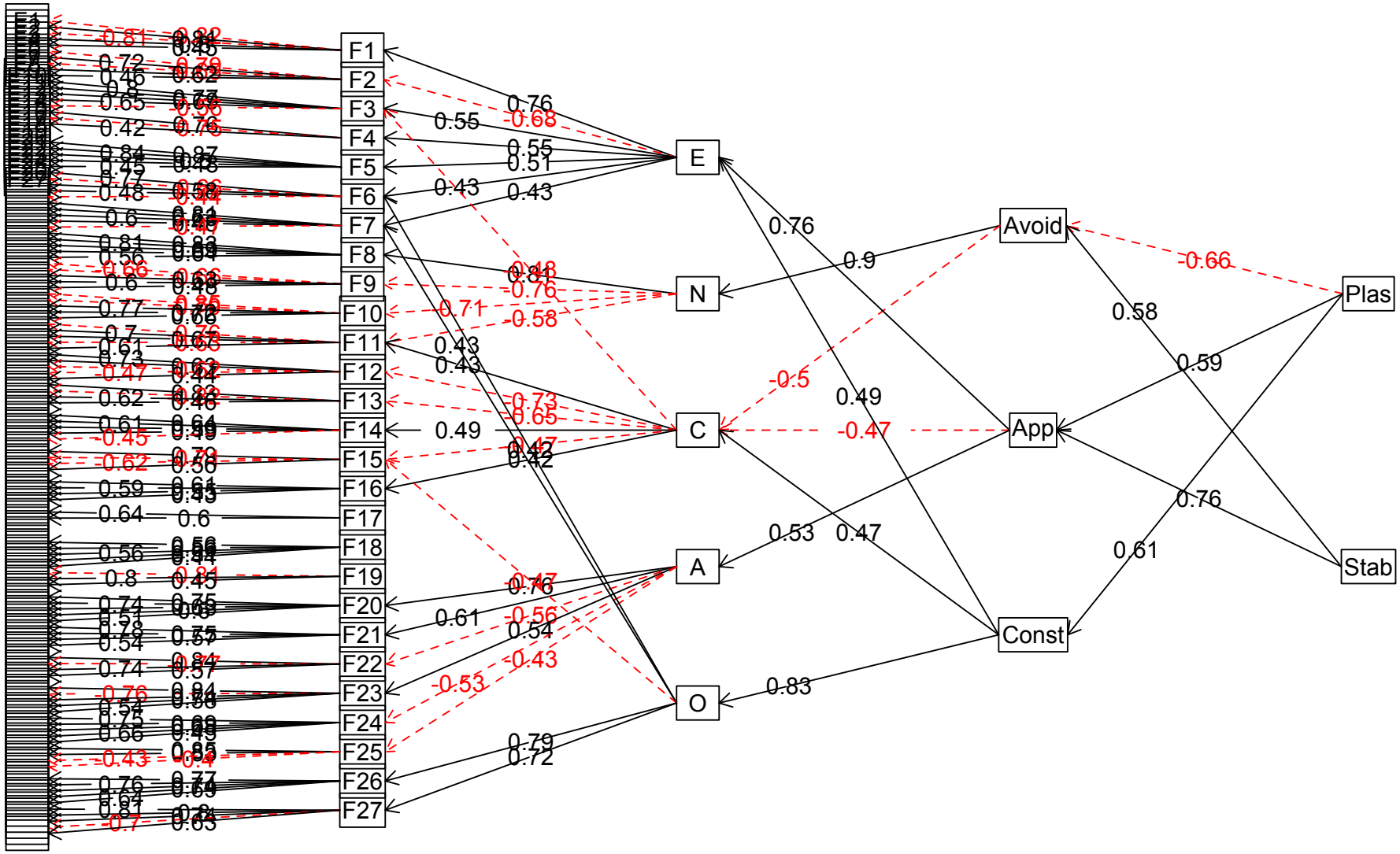
Prediction depends upon criteria: Goldberg: 72



Another factorial versus empirical example

1. SAPA Personality Inventory best 135 item (Condon (2018)
   - From 1800 IPIP items, found that 696 were most common
   - Factor structure of these 696 showed 135 very clear items
   - 5/27 factors, but not hiearchically organized
2. 4,000 subjects on spi 135 in the *psych* package
3. 135 items plus 10 criteria variables

Applying the 'Bass Ackward' function



Bass ackward is not a hierarchal solution

1. The “Bass-Ackward" algorithm (Goldberg, 2006; Waller, 2007) is a way of summarizing multiple solutions
2. Each solution is of the items
3. The factor scores (or their matrix equivalent) are then correlated
4. This is different from a hierarchal solution.

Score the Big 5 and predict the criteria

R code

```
spi.scales <- scoreItems(spi.keys[1:5],spi)

 cor2(spi[1:10],spi.scales$scores)
```

```
          Agree Consc Neuro Extra  Open
age        0.18  0.19 -0.17 -0.02  0.13
sex        0.17  0.09  0.24  0.06 -0.15
health     0.11  0.23 -0.34  0.21  0.07
p1edu      0.02 -0.02 -0.05  0.06  0.07
p2edu      0.02 -0.04 -0.04  0.08  0.07
education  0.13  0.12 -0.17 -0.01  0.15
wellness   0.11  0.12 -0.02  0.11  0.01
exer       0.07  0.19 -0.18  0.13  0.10
smoke     -0.09 -0.11  0.06  0.06  0.09
ER        -0.03 -0.01  0.12  0.02 -0.02
```

What about multiple R

R code

```
summary(setCor(1:10,11:15,data=spi.scores.df,plot=FALSE))
```

```
 summary(setCor(1:10,11:15,data=spi.scores.df,plot=FALSE))

Multiple Regression from raw data
setCor(y = 1:10, x = 11:15, data = spi.scores.df, plot = FALSE)

Multiple Regression from matrix input

Beta weights
        age    sex  health  p1edu  p2edu education wellness    exer  smoke     ER
Agree  0.16  0.162  0.0063  0.015  0.014     0.116   0.0631 -0.0053 -0.083 -0.025
Consc  0.13  0.103  0.1715 -0.034 -0.049     0.065   0.1053  0.1613 -0.082  0.016
Neuro -0.14  0.286 -0.2721 -0.036 -0.033    -0.147   0.0302 -0.1247  0.058  0.131
Extra -0.11  0.086  0.1436  0.047  0.061    -0.086   0.0918  0.0876  0.084  0.050
Open   0.12 -0.122  0.0126  0.058  0.057     0.142   0.0031  0.0675  0.090 -0.012

Multiple R
      age       sex    health     p1edu     p2edu education  wellness      exer     smoke
    0.306     0.360     0.405     0.098     0.109     0.264     0.170     0.267     0.181

Multiple R2
      age       sex    health     p1edu     p2edu education  wellness      exer     smoke
   0.0939    0.1296    0.1642    0.0096    0.0118    0.0699    0.0288    0.0711    0.0329

Cohen's set correlation R2
[1] 0.4

Squared Canonical Correlations
[1] 0.2394 0.1332 0.0620 0.0298 0.0079
```

Compare simple regression with mindless empiricism

1. Empirical scale construction (ala MMPI) can be done for any criterion
2. It ie essential to cross validate, for otherwise we are just over fitting (Cureton, 1950)
3. Traditional cross validation was splitting the sample in half, derive on one half, validate on the other half
4. Double cross validation was a simple improvement.
5. K-fold cross validation is a generalization of this procedure (k=2 is double cross validation).
6. Alternative is bootstrap over many (20-1000) alternatives.
7. Bagging is Boostrap aggregation
8. *bestscales* function (aka BISCUIT) will do this. (Elleman, McDougald, Revelle & Condon, 2020)

Compare with best scales

R code

```
bs <- bestScales(spi[11:145],spi[1:10],dictionary=spi.dictionary,n.ite
```

```
Call = bestScales(x = spi[11:145], criteria = spi[1:10], dictionary = spi.dictionary,
    n.iter = 20)
          derivation.mean derivation.sd validation.m validation.sd final.valid
age                  0.37         0.014        0.360         0.021        0.35
sex                  0.36         0.014        0.354         0.021        0.35
health               0.44         0.016        0.432         0.017        0.43
p1edu                0.15         0.030        0.124         0.026          NA
p2edu                0.17         0.027        0.098         0.024          NA
education            0.32         0.022        0.285         0.026        0.18
wellness             0.25         0.014        0.213         0.026        0.22
exer                 0.32         0.018        0.283         0.023        0.30
smoke                0.28         0.016        0.255         0.024        0.27
ER                   0.17         0.025        0.127         0.025        0.12

Repeat from setCor:
Multiple R
      age       sex    health     p1edu     p2edu education  wellness      exer     smoke
    0.306     0.360     0.405     0.098     0.109     0.264     0.170     0.267     0.181
```

What are the items?

Criterion = age

	Freq	mean.r	sd.r	item_id	item	item_scale	resp_type
q_4296	20	-0.23	0.01	q_4296	Tell a lot of lies.	EPQ:P	reg
q_4249	20	-0.21	0.02	q_4249	Would call myself a nervous person.	EPQ:N	reg
q_501	20	-0.21	0.01	q_501	Cheat to get ahead.	IPIP	reg
q_1024	18	-0.21	0.01	q_1024	Hang around doing nothing.	IPIP	reg
q_803	19	0.20	0.02	q_803	Express myself easily.	IPIP	reg
q_1081	18	-0.20	0.01	q_1081	Have difficulty expressing my feelings.	IPIP	reg

Criterion = sex

	Freq	mean.r	sd.r	item_id	item	item_s
q_1505	20	0.29	0.01	q_1505	Panic easily.	
q_979	20	0.29	0.01	q_979	Get overwhelmed by emotions.	
q_793	20	0.25	0.01	q_793	Experience my emotions intensely.	
q_174	20	-0.24	0.01	q_174	Am not easily affected by my emotions.	
q_1989	18	0.21	0.01	q_1989	Worry about things.	
q_851	19	0.21	0.01	q_851	Feel sympathy for those who are worse off than myself.	
q_1763	18	0.21	0.02	q_1763	Sympathize with others feelings.	
q_4252	18	0.20	0.01	q_4252	Am a worrier.	E

Criterion = health

	Freq	mean.r	sd.r	item_id	item	item_scale	resp
q_820	20	0.35	0.02	q_820	Feel comfortable with myself.	IPIP	
q_2765	20	0.35	0.01	q_2765	Am happy with my life.	IPIP	
q_811	20	-0.34	0.01	q_811	Feel a sense of worthlessness or hopelessness.	IPIP	
q_578	20	-0.34	0.02	q_578	Dislike myself.	IPIP	
q_1371	20	0.32	0.02	q_1371	Love life.	IPIP	
q_56	20	0.28	0.01	q_56	Am able to control my cravings.	IPIP	
q_1505	20	-0.27	0.01	q_1505	Panic easily.	IPIP	
q_808	18	-0.26	0.02	q_808	Fear for the worst.	IPIP	

Several classic and recent papers worth reading

1. Validity versus reliability (Loevinger, 1957; Steger, Jankowsky, Schroeders & Wilhelm, 2022)
2. The Great Response Style Myth (Block, 1965; Rorer, 1965)
   - Content dominates "yea saying" or social desirability
   - But perhaps extreme response style is a problem (Hamilton, 1968)
3. Number of alternatives (Simms, Zelazny, Williams & Bernstein, 2019)
4. The problem of detecting bad responders (Arias, Garrido, Jenaro, Martínez-Molina & Arias, 2020)
   - Is there a way to automate the detection of bad responders?
   - Does this make a difference?
   - Mixed model factor analysis as a way
   - Just examining inconsistencies between reversed items helps
5. Item wording effects positive versus negative wordings (Garcia-Pardina, Abad, Christensen, Golino & Garrido, 2022)

Validity versus reliability

(Adapted from Steger et al. (2022))

1. Reliability is more that α
2. Continuing debate about meaning of validity (Borsboom, 2006; Clark & Watson, 2019)
   - Measuring what a test purports to measure
   - Embedding in a nomological net (Cronbach & Meehl, 1955)
   - Ontology, reference and causality (Borsboom, Mellenbergh & van Heerden, 2004)
   - Prediction (Yarkoni & Westfall, 2017)
3. Three components of construct validity (Loevinger, 1957)
   - Substantive (the content)
   - Structural (factor structure)
   - External (convergent and discriminant)

Comparing approaches (Steger et al., 2022)

1. Three forms of scale construction
2. High reliability “ An item selection algorithm that focuses on factor saturation counteracts construct coverage, which resulted in scales that achieved high factor saturation, at the cost of being redundant in terms of content."
3. " In contrast, emphasizing construct coverage impedes factor saturation to a considerable extent, which also affects the interpretability of measurement models"
4. Balancing between the two using “Ant Colony Optimization"
5. For another demonstration of why high internal consistency does not enhance validity see Eagly & Revelle (2022)

Aggregation: effects on reliaiblity

For k standardized items with average correlations of $\bar{r}$, $\alpha = \frac{k\bar{r}}{1+(k-1)\bar{r}}$ or in terms of item variances (σ_i^2) and total test variance σ_x^2, and unknown error variance σ_e^2, the reliability, r_{xx} ,which is the expected correlation of a test with a test just like it is

$$r_{xx} = 1 - \frac{\sigma_e^2}{\sigma_x^2} = \alpha = \frac{k}{k-1}\frac{\sigma_x^2 - \Sigma(\sigma_i^2)}{\sigma_x^2}. \quad (3)$$

The square root of the reliability is the expected correlation with the domain that all of the items are supposedly measuring, and thus the upper bound of the test's validity. The reliability tends towards 1 as the the number of items increases.

Aggregation: validity

Less well known is the benefit of aggregation for predicting external criteria. If the average validity of an item is $\bar{r}_y$ and the average correlation within a composite remains $\bar{r}$, then the expected validity of a k item composite (r_{yc_k}) is just

$$r_{yc_k} = \frac{k\bar{r}_y}{\sigma_x} = \frac{k\bar{r}_y}{\sqrt{k + k * (k-1)\bar{r}}}. \quad (4)$$

That is, the sum of the individual validities divided by the square root of the variance of the composite. Clearly the aggregated validity increases with k and asymptotically tends towards $r_{cy_\infty} = \frac{\bar{r}_y}{\sqrt{\bar{r}}}$. For a fixed average item validity, test validity is a positive function of the number of items and is higher the lower the correlations between the items within the composite. The same features that increase reliability ($\bar{r}$) for a composite (Equation 3) decrease the asymptotic validity r_{yc_∞}.

An example from the Athenstaedt data

We show this with an examination of scales of length 5, 10, 20, 30, 40, 50, 60, and 70 chosen from the Athenstaedt (2003) data set using the `bestScales` function from the *psych* package (Revelle, 2022). As we increase the number of items in the scale, the average validity of the items decrease, as does the average correlation of the predictor set. But the validity increases.

Table: Choosing the best k items to predict sex in the Athenstaedt (2003) data set. r is the correlation of a k-item scale with sex, avrg is the average correlation with the predictor set, alpha is the alpha reliability of the predictor set. Means show the average validity of the items used in the scale.

A table from the psych package in R

k	scale validity	avrg.	alpha	mean item validity
5	0.66	0.14	0.49	0.43
10	0.74	0.13	0.62	0.40
20	0.77	0.11	0.72	0.35
30	0.76	0.10	0.77	0.32
40	0.76	0.09	0.80	0.29
50	0.75	0.08	0.81	0.26
60	0.75	0.06	0.81	0.24
70	0.72	0.05	0.79	0.21

Another data set

I show these relationships in (Table 2) and a graphic (Figure 1).

Table: Exploring the benefits and costs of aggregation. Although reliability will increase, because the items were chosen in order of their validity, scale validity is non-monotonic with the number of items (see figure). The ratio is just the average validity/sqrt(average item correlation.

Reliability and validity of various length scales when items are chosen by their validity.

Variable	N.items	alpha	validity	average.r	item.validity	ratio	modeled
five	5	0.67	0.65	0.29	0.43	0.80	0.65
ten	10	0.76	0.71	0.24	0.40	0.82	0.71
fifteen	15	0.80	0.73	0.21	0.37	0.81	0.73
twenty	20	0.82	0.74	0.19	0.35	0.82	0.74
thirty	30	0.85	0.73	0.16	0.32	0.79	0.73
fourty	40	0.87	0.73	0.14	0.29	0.78	0.73
fifty	50	0.88	0.72	0.13	0.27	0.77	0.72
fiftysix	56	0.89	0.70	0.12	0.26	0.75	0.70

The power of aggregation

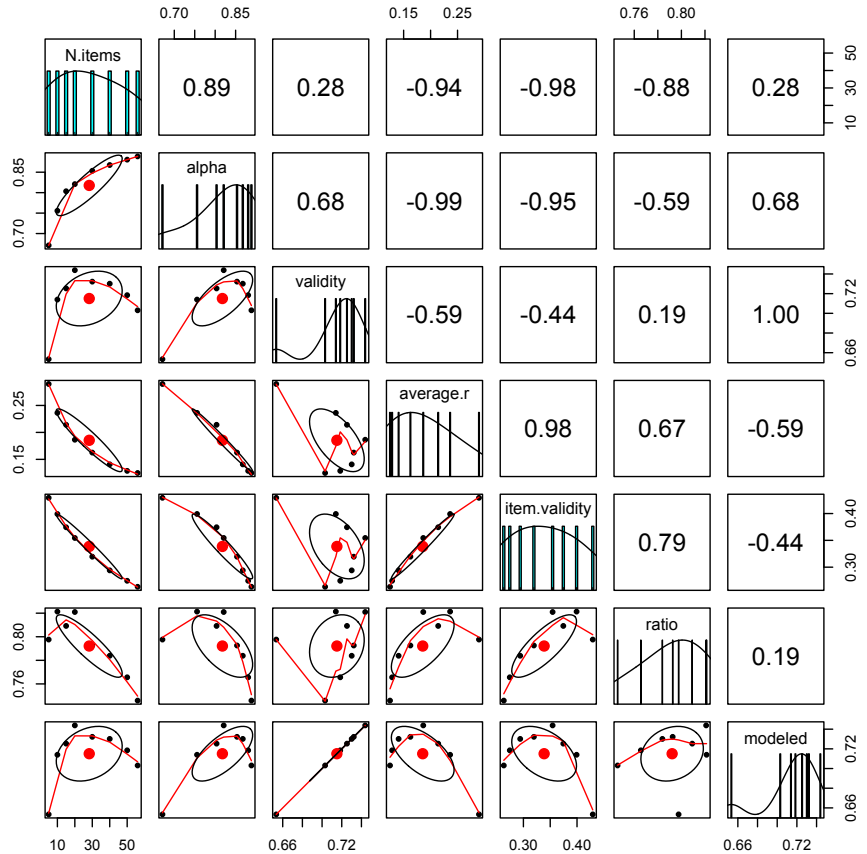


Figure: If the items are chosen based upon their validities, reliability of a scale increases with number of items, but validity is a non-monotonic function of the number of items. This is because we are using the best items first.

Randomly choose items (from the domain)

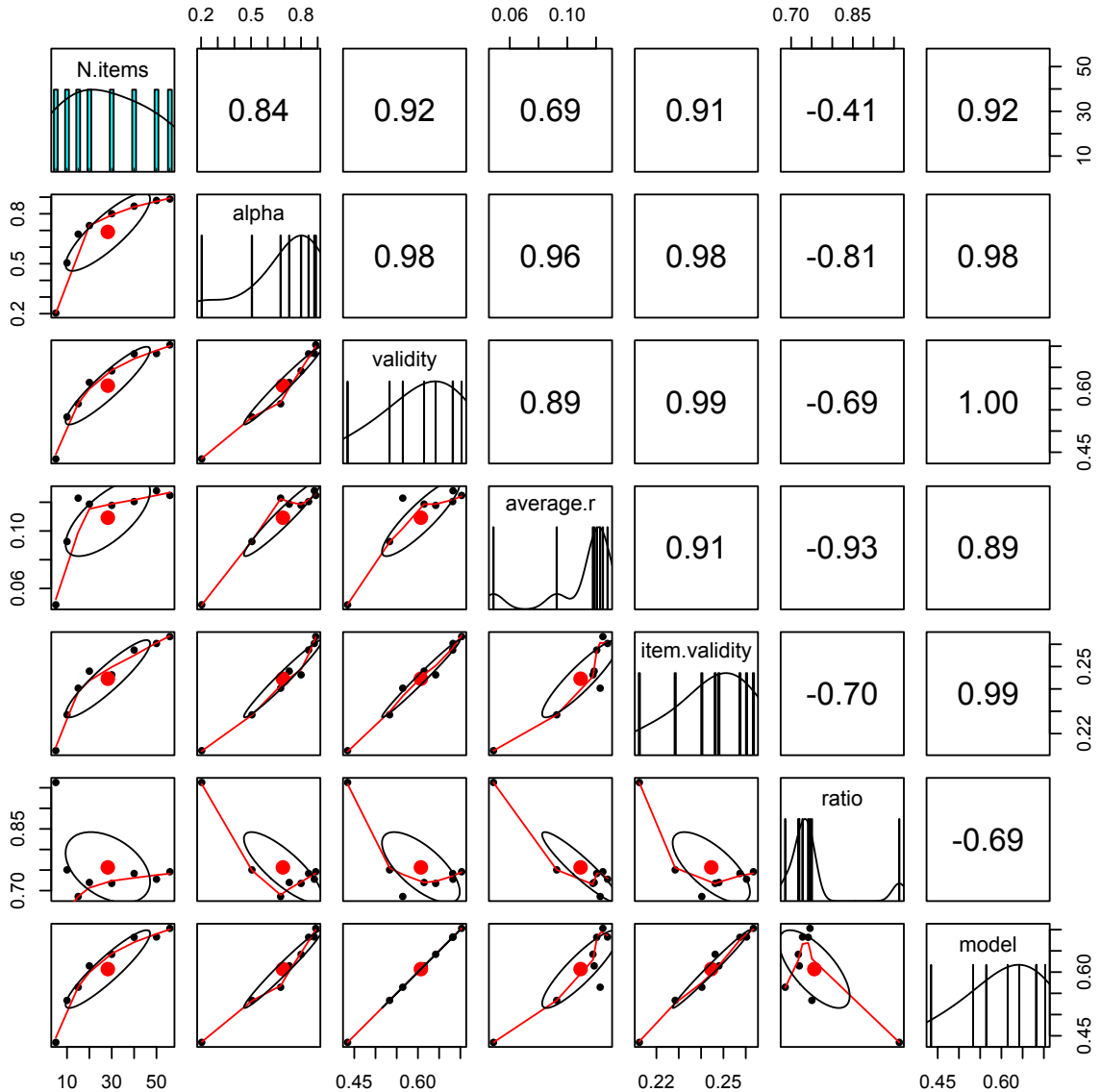
The prior analysis was choosing items in terms of their validity. That is to say, we take the cream first. Compare this to just randomly choosing items. In this case, as the number of items being aggregated increases, the validity increases as predicted by Equation 4.

Table: The item and scale statistics when scales are formed from random subsets of domain items. The ratio is just the average validity/sqrt(average item correlation).

Reliability and validity of various length scales when items are chosen randomly.

Variable	N.items	alpha	validity	average.r	item.validity	ratio	modeled
r.five	5	0.20	0.43	0.05	0.21	0.96	0.43
r.ten	10	0.51	0.53	0.09	0.23	0.75	0.53
r15	15	0.68	0.56	0.12	0.24	0.69	0.56
r20	20	0.73	0.61	0.12	0.25	0.72	0.61
r30	30	0.80	0.64	0.12	0.25	0.72	0.64
r40	40	0.85	0.68	0.12	0.26	0.74	0.68
r50	50	0.88	0.68	0.13	0.26	0.73	0.68
all.56	56	0.89	0.70	0.12	0.26	0.75	0.70

Randomly choose items (from the domain)



Summary of scale construction

1. Define the domain of interest
2. Create items to assess that domain
3. Examine the internal structure of the measure
4. Include supposedly unrelated items (hyperplane stuff)
5. Worry about response characteristics
6. Consider the purpose of the scale (measuring a domain, predicting some criterion)

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, *52*(6), 2489–2505.

Athenstaedt, U. (2003). On the content and structure of the gender role self-concept: Including gender-stereotypical behaviors in addition to traits. *Psychology of Women Quarterly*, *27*(4), 309–318.

Block, J. (1965). *The challenge of response sets: Unconfounding meaning, acquiescence, and social desirability in the MMPI*. East Norwalk, CT: Appleton-Century-Crofts.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425–440.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071.

Clark, L. A. & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological assessment*, *31*(12), 1412.

Condon, D. M. (2018). The SAPA Personality Inventory: An empirically-derived, hierarchically-organized self-report personality assessment model. *PsyArXiv*.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.

Cureton, E. E. (1950). Validity, reliability, and baloney. *Educational and Psychological Measurement*, *10*(1), 94–96.

Eagly, A. H. & Revelle, W. (2022). Understanding the Magnitude of Psychological Differences between women and men requires seeing the forest and the trees (in press). *Perspectives in Psychological Science in press*.

Elleman, L. G., McDougald, S., Revelle, W., & Condon, D. (2020). That takes the BISCUIT: a comparative study of predictive accuracy and parsimony of four statistical learning techniques in personality data, with data missingness conditions. *European Journal of Psychological Assessment*, *36*(6), 948–958.

Garcia-Pardina, A., Abad, F. J., Christensen, A. P., Golino, H., & Garrido, L. E. (2022). Dimensionality assessment in the

presence of wording effects: A network psychometric and factorial approach. *PsyArXiv PsyArXiv*.

Goldberg, L. R. (1972). Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monographs. No 72-2*, *7*.

Goldberg, L. R. (2006). Doing it all bass-ackwards: The development of hierarchical factor structures from the top down. *Journal of Research in Personality*, *40*(4), 347 – 358.

Hamilton, D. L. (1968). Personality attributes associated with extreme response style. *Psychological bulletin*, *69*(3), 192.

Hase, H. D. & Goldberg, L. R. (1967). Comparative validity of different strategies of constructing personality inventory scales. *Psychological Bulletin*, *67*(4), 231–248.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports Monograph Supplement 9*, *3*, 635–694.

Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research* (2.2.3 ed.). psych: Northwestern University, Evanston. R package version 2.2.3.

Revelle, W., Dworak, E. M., & Condon, D. M. (2021). Exploring the persome: The power of the item in understanding personality structure. *Personality and Individual Differences*, *169*.

Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, *63*(3), 129 – 156.

Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? psychometric perspectives using personality questionnaire data. *Psychological Assessment*, *31*(4), 557–566.

Steger, D., Jankowsky, K., Schroeders, U., & Wilhelm, O. (2022). The road to hell is paved with good intentions: How common practices in scale construction hurt validity. *PsyArXiv*.

Waller, N. (2007). A general method for computing hierarchical component structures by Goldberg's Bass-Ackwards method. *Journal of Research in Personality*, *41*(4), 745 – 752.

Yarkoni, T. & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122.