Using MMCAR to explore the structure of personality and ability

Part of an invited symposium (chaired by Jingchen Liu): Big data analysis in cognitive assessment and latent variable models International Meeting of Psychometric Society, Zurich,

William Revelle[a] & David M. Condon[b]

[a]Department of Psychology, Northwestern University, Evanston, Illinois
[b]Department of Medical Social Sciences, Northwestern University Chicago Illinois
Partially supported by a grant from the National Science Foundation: SMA-1419324
Slides at http://personality-project.org/sapa.html

NORTHWESTERN UNIVERSITY

Outline

Abstract

Personality and ability item pools can be very large ($> 5,000$) but no one person likes to answer more than 50-150 items. Using web based "Synthetic Aperture Personality Assessment" at sapa-project.org we collect data using a Massively Missing Completely at Random (MMCAR) design.
1,000 x 1,000 covariance matrices based upon 250,000 subjects using pairwise covariances have roughly 2 - 3,000 observations/pair. Using conventional covariance algebra and R functions in the psych package, we can examine the joint structure of personality, ability, and interests.The advantages of a SAPA/MMCAR design compared to the more conventional use of short forms will be discussed. In particular, by using large samples with many items, the structure of personality can be examined at many levels of resolution, from the conventional 3-5 factors to our preferred 27 homogeneous scales, down to the unique (but stable) correlates of individual items.

The basic problem: Fidelity versus bandwidth

1. Many personality traits, interests and cognitive abilities are multidimensional and have complex structure.
    - To measure these, we need to have the precision that comes with many participants.
    - But we also need the bandwidth that comes with many items.
    - But participants are reluctant to answer very many items.
2. This has led to the quandary of should you give many people a few items or a few people, many items?
3. Our answer is to do both, but with a *Massively Missing Completely At Random* (MMCAR) data structure.
4. We refer to this technique as *Synthetic Aperture Personality Assessment* (SAPA) to recognize the analogy to synthetic aperture radio astronomy (Revelle, Wilt & Rosenthal, 2010; Revelle, Condon, Wilt, French, Brown & Elleman, 2016)
5. This is functionally what Frederic Lord (1955, 1977) suggested 62 years ago. It is time to take him seriously.

Breadth vs. depth of measurement

1. Factor structure of domains needs multiple constructs to define structure.
2. Each construct needs multiple items to be measured reliably.
3. This leads to an explosion of potential items.
4. But, people are willing to only answer a limited number of items.
5. This leads to the use of short and shorter forms (the NEO-PI-R (Costa & McCrae, 1992) with 300, the IPIP (Goldberg, 1999) Big 5 with 100, the BFI (John, Donahue & Kentle, 1991) with 44 items, the BFI2 (Soto & John, 2017) with 60, the TIPI (Gosling, Rentfrow & Swann, 2003) with 10 and 10 item BFI (Rammstedt & John, 2007)) to include as part of other surveys.
6. Unfortunately, with this reduction of items, breadth of substantive content is lost.

Example studies with subject/item tradeoffs

1. The Potter-Gosling internet project (outofservice.com) has given over 10,000,000 tests since 1997. Originally the 44 items of the Big Five Inventory (BFI) (John et al., 1991) although they are now giving the BFI2 (Soto & John, 2017)
2. The Stillwell-Kosinski (mypersonality.org) Facebook application (no longer in service) gave 7,765 people the IPIP version of the NEO-PI-R with facets (300 items), 1,108,472 the IPIP NEO-PI R domains (100 items), and 3,646,237 brief (20 item) surveys. Cross linked to likes and Facebook pages (Kosinski, Matz, Gosling, Popov & Stillwell, 2015; Youyou, Kosinski & Stillwell, 2015)
3. Smaller scale studies include the initial report on the BFI-2 (Soto & John, 2017) with several thousand subjects with 60 item.

Exceptions to the shorter and shorter inventory trend

1. Lew Goldberg and his colleagues at the University of Oregon developed the Eugene-Springfield sample (Goldberg & Saucier, 2016) which has given several thousand items to $\approx 1,000$ participants over 10 years. This sample has been the basis of the development and validation of the International Personality Item Pool (see ipip.ori.org). In fact, many of the subsequent attempts at personality scale development have used the Eugene-Springfield sample, e.g., the BFI (John et al., 1991), and the Big Five Aspect Scales (BFAS) of DeYoung, Quilty & Peterson (2007).
2. Our Personality Project (Revelle et al., 2010, 2016) (now at sapa.project.org) has taken the opposite direction and has given more and more items including measures of temperament, ability, and interests and we are now developing item statistics on more than 4,000 items (Condon & Revelle, 2017) for more than 250,000 participants (but uses SAPA procedures).

Trading items for people: Studies, Items, People, Items x People

Table: Data sets vary in their sampling strategy and the Potter-Gosling and Stillwell/Kosinski data sets seem to have more data than the others

Study	N	Items (k)	Items/ Person	Items* People
Potter-Gosling	10^7	44	44	$4.4 * 10^8$
Stillwell-Kosinski	$4.5 * 10^6$	20-300	20-300	$1.7 * 10^8$
SAPA	$2.5 * 10^5$	1-4,000	100-150	$2.5 * 10^7$
Eugene-Springfield	10^3	3,000	3,000	$3 * 10^6$

But given basic statistical theory, is it worth while to increase the sample size so much? What is the effect of giving more items at the cost of reducing the sample size?

Consider the amount of *information* which varies by number of correlations $\frac{k*(k-1)}{2}$ and $\sqrt{N}$.

Trading items for people: Studies: Items, People, Items x People and Information

Information varies by the number of correlations $(k * (k - 1)/2)$ weighted by their standard errors which vary by $\sqrt{N}$

Table: Data sets vary in their sampling strategy and the seemingly smaller sets, by giving many more items actually have more total information

Study	N	Items	Items/ Person	Items* People	Information
SAPA*	$2.5 * 10^5$	1-2,000	100-150	$2.5 * 10^7$	$2.5 * 10^8$
E-S	1,000	3,000?	3,000	$3 * 10^6$	$1.4 * 10^8$
S-Ki	$4.5 * 10^6$	20-300	20-300	$1.7 * 10^8$	$9.5 * 10^7$
SAPA*	$4.3 * 10^3$	953	100-150	$2.5 * 10^7$	$3 * 10^7$
P-G	10^7	44	44	$4.4 * 10^8$	$3.0 * 10^6$

*The average pairwise count of observations for the SAPA data reported today are 4,291 from 250,000 total participants for 953 items.

Many items versus many people

1. Not only do want many people, we also want many items.
2. Resolution (fidelity) goes up with sample size, N, (standard errors are a function of $\sqrt{N}$)

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N-1}} \qquad \sigma_r = \frac{1-r^2}{\sqrt{N-2}}$$

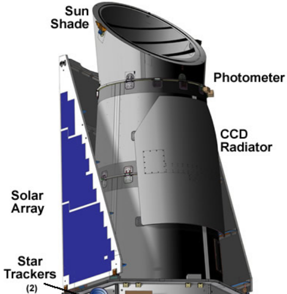
3. Also increases as number of items, k, measuring each construct (reliability as well as signal/noise ratio varies as number of items and average correlation of the items)

$$\lambda_3 = \alpha = \frac{k\bar{r}}{1+(k-1)\bar{r}} \qquad s/n = \frac{k\bar{r}}{(1-k\bar{r})}$$

4. Breadth of constructs (band width) measured goes up by number of items (k).
5. Thus, we need to increase N as well as k. But how?

A short diversion: the history of optical telescopes

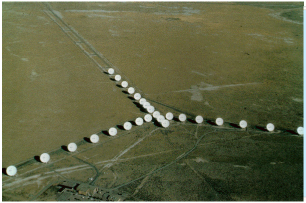
Resolution varies by aperture diameter (bigger is better)

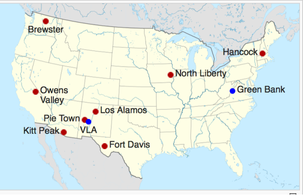


A short diversion: history of radio telescopes

Resolution varies by aperture diameter (bigger is still better)

Aperture can be synthetically increased across multiple telescopes or even multiple observatories





Can we increase N and n at the same time?

1. Frederic Lord (1955) introduced the concept of sampling people as well as items.
2. Apply basic sampling theory to include not just people (well known) but also to sample items within a domain (less well known).
3. Basic principle of Item Response Theory and tailored tests.
4. Used by Educational Testing Service (ETS) to pilot items.
5. Used by Programme for International Student Assessment (PISA) in incomplete block design (Anderson, Lin, Treagust, Ross & Yore, 2007).
6. Discussed at this meeting by Rutkowski and Matta in the missing data symposium.
7. Can we use this procedure for the study of individual differences without being a large company?
8. Yes, apply the techniques of radio astronomy to combine measures synthetically and take advantage of the web.

Subjects are expensive, so are items

1. In a survey such as Amazon's Mechanical Turk (MTURK), we would need to pay by the person and by the item.
2. Volunteer subjects are not very willing to answer many items.
3. Why give each person the same items? Sample items, as we sample people.
4. Synthetically combine data across subjects and across items. This will imply a missing data structure which is
   - Missing Completely At Random (MCAR), or even more descriptively:
   - Massively Missing Completely at Random (MMCAR) (we sometimes have 99% missing data although our median is only 93% missing!)
5. This is the essence of Synthetic Aperture Personality Assessment (SAPA) (Condon & Revelle, 2014; Condon, 2014; Revelle et al., 2016, 2010).
6. This is a much higher rate of missingness than discussed in the balanced incomplete block design of NAEPS or PISA.

3 Methods of collecting 256 subject * items data

a) 8 x 32 complete	b) 32 x 8 complete	c) 32 x 32 MCAR p=.25
46213634521143453443645331212414	46323114	..3..2..6.....4.55.......44.....
21243623166421516154432261516513	25443314	...........4..6..45..3.4..6....1
51661351155165463622224435623344	43315423	6..3.......6.1.....6.2.......5.6
11141343362332215612152135614522	26314145	....3522......5.3...3......5....
25353121264561433433232246526411	41435614	....3.2.2.......3..2......65..5.
61335154566424114612641225353516	42236153	.....51....324.........23......5
24634342151536242425413513435116	62421344	....552............25...54.5....
11554654453123111162423325516334	35234443	...44.4.5....3..6...6........3..
	34514166	....61.523.2....2...........3...
	63415154	5.............42.4..6.5......61.
	44441342	....3....3.6..1.4...1..5......5.
	13514321	1....54..........2.4.33..6......
	66365663	4.....52..6.....44.3...........2
	12264546	..44...1........1..42....5..1...
	31466135	..1..3.......2..3.521.......6...
	32645514	.......3.142.........22......12.
	66151251	.4...2..........3..162...4.....4
	14411441	..4..6..3.4...1....5.33.........
	62443636	5..........243..5....41......1..
	33316236	..5..3..4...4.4..5..1.........4.
	63325425	.....4.......3..5.2.....64.4..4.
	11531126	...1.1.2...6....4......55....2..
	61155546	.....3..2..53.....2..2.3.3......
	33245361	......1...2..43...3.13........5.
	52241654	...2.........4..54...2.3..62....
	63212356	22.......332..1.....5......6....
	24414663	...5..3.4.....3....5.241........
	63661414	......63.1.......6...5..4..2...5
	45555223	..2.4..5..........52.4.....44...
	14364433	2.55.....2.....6.....6.....55...
	21461416	..5..........4....6341.4..2.....
	33232365	....55......5.......45....3..32.

Synthetic Aperture Personality Assessment

1. Give each participant a random sample of *pn* items taken from a larger pool of n items. p_i might be anywhere from .01 to 1.
2. Find covariances based upon "pairwise complete data". Each pair appears with probability $p_i p_j$ with a median of .01.
3. Find scales based upon basic covariance algebra.
   - Let the raw data be the matrix ${}_N\boldsymbol{X}_n$ with N observations converted to deviation scores.
   - Then the item variance covariance matrix is ${}_n\boldsymbol{C}_n = \boldsymbol{X'X}N^{-1}$
   - and scale scores, ${}_N\boldsymbol{S}_s$ are found by $\boldsymbol{S} = {}_N\boldsymbol{X}_{pp}\boldsymbol{K}_s$.
   - ${}_n\boldsymbol{K}_s$ is a keying matrix, with $k_{ij} = 1$ if $item_i$ is to be scored in the positive direction for scale j, 0 if it is not to be scored, and -1 if it is to be scored in the negative direction.
   - In this case, the covariance between scales, ${}_s\boldsymbol{C}_s = {}_s\boldsymbol{S}_N{}'{}_N\boldsymbol{S}_s N^{-1} =$

$$ {}_s\boldsymbol{C}_s = (\boldsymbol{XK})'(\boldsymbol{XK})N^{-1} = \boldsymbol{K'X'XK}N^{-1} = \boldsymbol{K'}{}_n\boldsymbol{C}_n\boldsymbol{K}. \quad (1) $$

4. That is, we can find the correlations/covariances between scales from the item covariances, not the raw items.

Two sets of simulations

1. For both sets, we consider the effect of sample size (N) and sampling probability (p).
2. The effect on standard errors of intercorrelations of item composites of composite length
3. Examining the standard errors of factor loadings
4. We consider effective sample size and compare it to the nominal sample size of pairwise correlations.

The basic tradeoff: standard errors and effective sample size

1. Standard error of correlations between any *single pair of items* is just
$$\sigma_r = \frac{1 - r^2}{\sqrt{N-2}}$$
2. However, simulation (and some theory) shows that the standard error of correlations of *synthetic* correlations of *scales of length k* decreases as a joint function of the number of items in the scale and the *inverse of the probability* of any two items being administered.
3. Effectively, this is because what ever causes error in any correlation does not aggregate across k independent pairs of items.

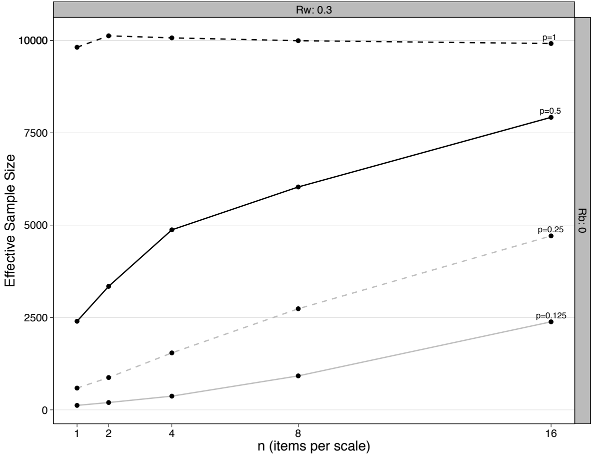
SAPA standard errors of correlations vary by scale length

1. When forming synthetic scales from MMCAR based items, the standard error of correlations decreases as a function of the Total number of subjects (N), the *the inverse of the percentage* of items sampled (p), and the *number of items forming the scale* (k).
2. Ashley Brown has shown this quite clearly by simulation (Brown, 2014) and we discussed this last year at APS (Revelle & Condon, 2016) and in a recent chapter (Revelle et al., 2016),
3. A good way to visualize this is to examine the standard error of correlations as a function of N, p, and k.
4. An even more dramatic way is to plot the *Effective Sample Size* (N_{eff}) which because

$$\sigma_r = \frac{1 - r^2}{\sqrt{N - 2}} \text{ is merely } \quad N_{eff} = \frac{(1 - r^2)^2}{\sigma_r^2} + 2$$

Effective sample size varies by the size of the composite scale.

Simulating N= 10,000 with probability of any item (Brown, 2014) ($p = .125, .25, .5$, or 1) and items in the composite 1 , 2, 4, 8, 16.



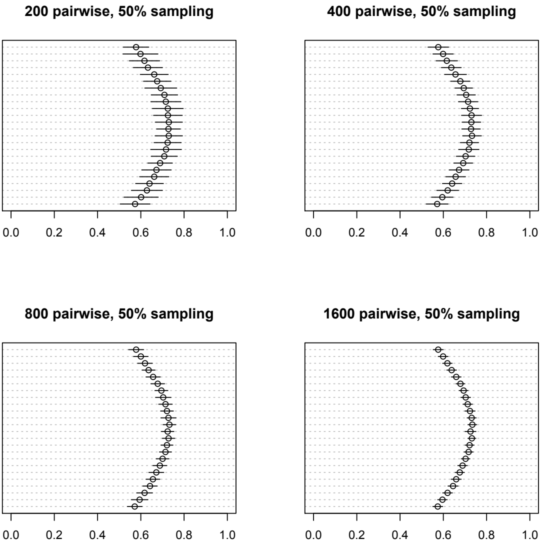
Comments on simulation values

1. These simulations are based upon N = 10,000
2. Although for k = 1, the effective sample size is, of course, just Np^2 and thus for $p = .25 = 10,000 * .25^2 = 625$ this provides a relatively small standard error ($\sigma_r = .04$).
3. Had we not sampled, we would have a standard error of .01 but for 1/4 the number of items and thus 1/16 the number of correlations.
4. Is this extra precision worth the reduction in bandwidth?
5. More importantly, the standard error of 4 items scales with an even more dramatic sampling (p = .125) would also be roughly .04 but with 8 times as many items and thus 64 times as many correlations.

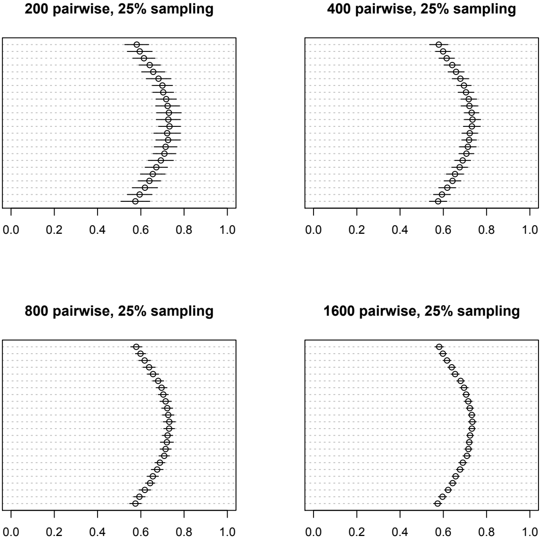
Standard deviations of factor loadings

1. 3 Simulation of SAPA procedures
2. Generate 200,000 simulated cases with a 1 factor model
3. Compare no sampling, 50% sample, and 25% sample
4. This leads to 100%, 25% or 6.25% pairwise correlations.
5. The question becomes what is the effective size? Is it the number of pairwise observations or something greater?

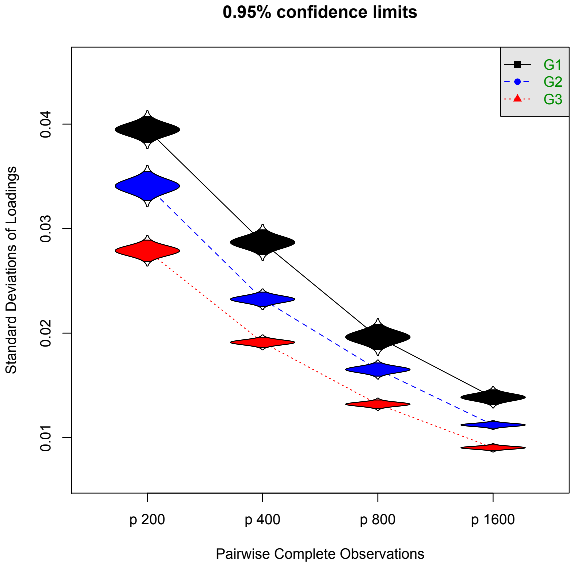
Factor loadings and standard errors for 50% sampling



Factor loadings and standard errors for 25% sampling

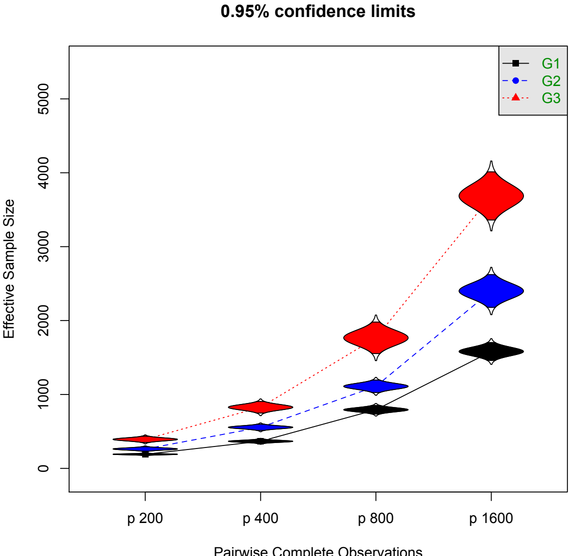


Standard deviations of factor loadings



1. "Normal " => pairwise = sample size
2. 50% sample => pairwise = .25 of sample size
3. 25% sample = .0625% pairwise

Effective sample size based upon sampling variability of factor loadings



1. 25% sample => total sample = 16 times pairwise sample
2. 50% sample => total sample = 4 times pairwise sample
3. "Normal " => pairwise = sample size

Conclusions based upon simulations

1. Massively missing completely at random (MMCAR/SAPA) procedures increase bandwidth and fidelity.
2. For the same number items * people, sampling items increases the number of items that can be studied
3. Precision of scale x scale correlations and factor loadings within a scale is increased by sampling items
4. However, to achieve stable number of pairwise correlations, we need to increase overall number of subjects studied.
5. Nice in theory, does it work in practice? Yes, the SAPA project.

Integrating measures of Temperament , Ability and Interests (TAI)

1. Temperament
   - 696 items taken from 200 International Personality Item Pool (IPIP) (Goldberg, 1999) scales (representing 2,084 overlapping items)
   - 100% coverage of 200 IPIP scales, 57% to 85% of 235 other scales. (Condon, 2014, 2017; Condon & Revelle, 2017)
2. Ability
   - 60 items from the original International Cognitive Ability Resource (ICAR) data set (Condon & Revelle, 2014)
   - The original 4 item types: Verbal Reasoning, Matrix Reasoning, Number/Letter series, and 3 Dimensional rotation,
   - Now supplemented with 12 more item types as part of the ICAR project done in collaboration with Phillip Doebler, Heinz Hollng and Ehsan Masoudi in Germany; John Rust, David Stillwell, Luning Sun, Fiona Chan and Aiden Loe from the UK.
   - See http://ICAR-project.com for more information
3. Interest
   - Items taken from the Oregon Vocational Interests (ORVIS) (Pozzebon, Visser, Ashton, Lee & Goldberg, 2010) and Occupational Interests (ORAIS) (Goldberg, 2010). See Elleman, Condon & Revelle (2017).

How stable are factor analytic solutions of SAPA data?

1. Compare solutions for entire sample (255K) to subsamples
2. Examine the identification of structure of the ICAR16 sample test for a 4 factor solution
3. Examine the stability of factor loadings for the 3 Dimensional Rotation items (relevant for doing subsequent IRT scoring)
4. Show variation in loadings and in interfactor correlations for 100 replications of 5, 10, 20, and 40 % samples.
5. Compare hierarchical (omega) factor solution for entire sample to 2 and 5% samples.
6. Examine the stability of factor solutions for simulated data varying the probability of item selection.

Sample characteristics

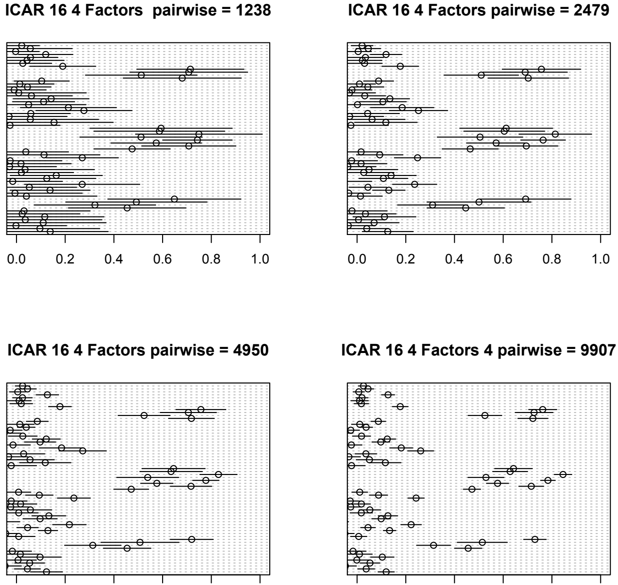
Items presented are sample from subsets of items (e.g. there are 24 3D rotation items and thus the probability of any pair being presented is less than the 16 sample test (ICAR 16) items.

Table: Sample sizes and number of times items were presented as well as pairwise counts

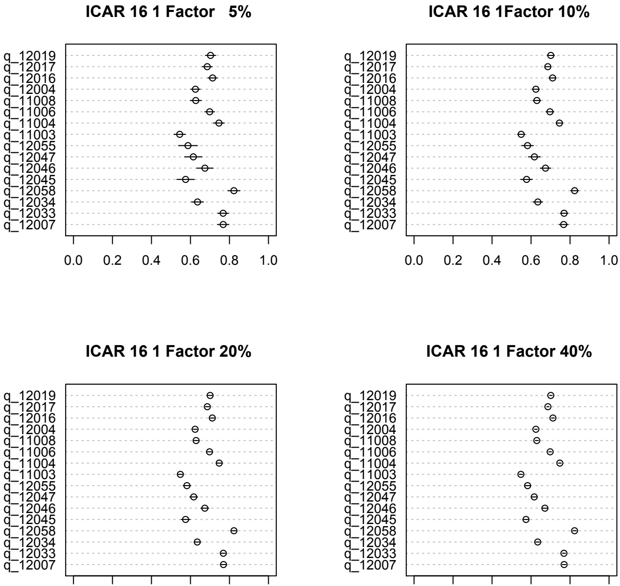
Fraction	sample	icar16	ic16 pairs	3DRot	3D pairs
0.05	12,767	3,717	1,238	1,668	200
0.10	25,535	7,435	2,479	3,335	399
0.20	51,070	14,869	4,950	6,670	798
0.40	102,139	29,738	9,907	13,340	1,596
1.00	255,348	74,345	24,768	33,351	3,990

We want to examine the stability of solutions across sampling frames.

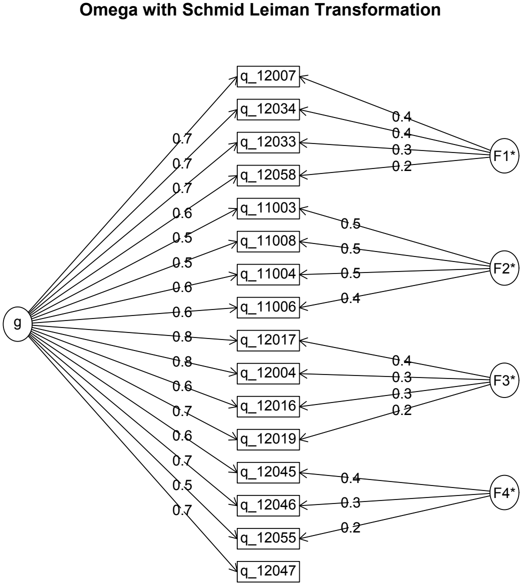
Stability of factor loadings for a 4 factor solution – Summary



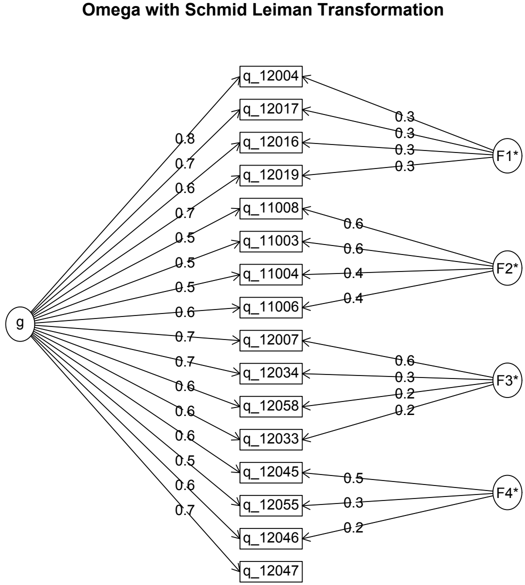
Stability of factor loadings for a 1 factor solution – Summary



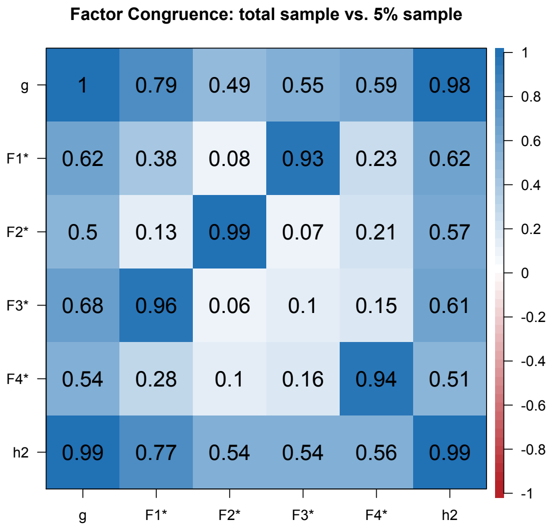
Hierarchical factor analysis – omega analysis complete sample



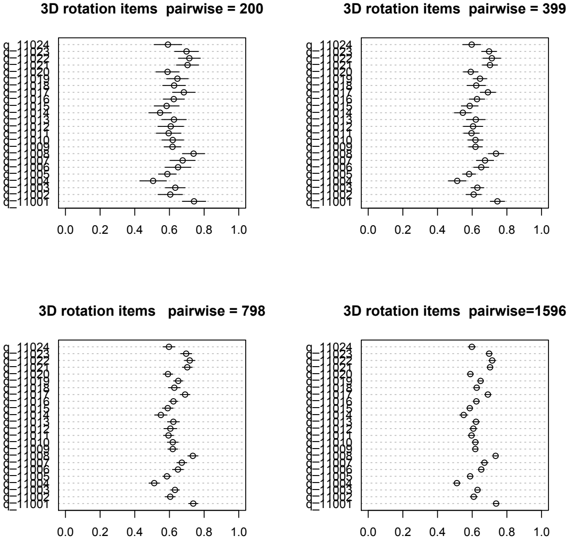
Hierarchical factor analysis – omega analysis 5% sample



Factor congruence of omega solutions: complete vs. 5% sample



Factor loadings for the 24 3D rotation items: summary



SAPA or MMCAR procedures are very powerful

1. Able to estimate difficulty parameters and covariance structures for 1,000s of items even though only 100 items answered per subject
2. Structure of ability measures using the open source ability test from the International Cognitive Ability Resource (ICAR) http://icar-project.com
3. Join the ICAR and SAPA projects.
4. Data sharing: https://dataverse.harvard.edu/dataverse/SAPA-ProjectCode/manuscript/
5. workbook: https://sapa-project.org/research/SPI/SPIdevelopment.pdf
6. SPI scales, norms, IRT parameters: https://sapa-project.org/research/SPI
7. Today's slides at http://personality-project.org/sapa.html

R code

The next few slides show the R code used for these analyses

R code

```
sapa <- read.file()  #this searched for and loaded the master data file
icar.dictionary <- read.file()  #this searches for and loads a dictionary file

sapa <- SAPAdata18aug2010thru7feb2017  #change the name to make the following analyses cleare
 #the raw sapa has alphanumeric codes for some fields. Convert to numeric
sapa <- char2numeric(sapa)
 icar <- sapa[rownames(icar.dictionary)] #Identify the ICAR items
 icar.keys <- ItemLists[417:421]  #the keys are loaded as part of the sapa read.file line
icar  #show the item numbers of ICAR

icar.16 <- c("q_12007","q_12033" ,"q_12034","q_12058","q_12045", "q_12046","q_12047",
"q_12055", "q_11003", "q_11004","q_11006" ,"q_11008","q_12004","q_12016","
q_12017", "q_12019")

#now do random resamples, 100 times each of various fractions of the icar.16 data
 icar.16.sapa.05 <-fa.sapa(sapa[icar.16],4,n.iter=100,frac=.05)
  icar.16.sapa.1 <-fa.sapa(sapa[icar.16],4,n.iter=100,frac=.1)
 icar.16.sapa.2  <-fa.sapa(sapa[icar.16],4,n.iter=100,frac=.2)
 icar.16.sapa.4  <-fa.sapa(sapa[icar.16],4,n.iter=100,frac=.4)
```

More R

R code

```
error.dots(icar.16.sapa.05,head=40,tail=40,sort=FALSE,main="ICAR 16 4 Factors  pairwise = 123
error.dots(icar.16.sapa.1,head=40,tail=40,sort=FALSE,main="ICAR 16 4 Factors pairwise = 2479"
error.dots(icar.16.sapa.2,head=40,tail=40,sort=FALSE,main="ICAR 16 4 Factors pairwise = 4950"
 error.dots(icar.16.sapa.4,head=40,tail=40,sort=FALSE,main="ICAR 16 4 Factors 4 pairwise = 99


 #now show the factor intercorrelations
 names(icar.16.sapa.05$cis$means.rot) <- label.icar.rot
names(icar.16.sapa.1$cis$means.rot) <- label.icar.rot
 names(icar.16.sapa.2$cis$means.rot) <- label.icar.rot
 names(icar.16.sapa.4$cis$means.rot) <- label.icar.rot

  error.dots(icar.16.sapa.05$cis$means.rot,se=icar.16.sapa.05$cis$sds.rot,sort=FALSE,xlim=c(0
  error.dots(icar.16.sapa.1$cis$means.rot,se=icar.16.sapa.1$cis$sds.rot,sort=FALSE,xlim=c(0,1
  error.dots(icar.16.sapa.2$cis$means.rot,se=icar.16.sapa.2$cis$sds.rot,sort=FALSE,xlim=c(0,1
  error.dots(icar.16.sapa.4$cis$means.rot,se=icar.16.sapa.4$cis$sds.rot,sort=FALSE,xlim=c(0,1

  op <- par(mfrow=c(2,2))
   error.dots(icar.16.sapa.05$cis$means.rot,se=icar.16.sapa.05$cis$sds.rot,sort=FALSE,xlim=c
  error.dots(icar.16.sapa.1$cis$means.rot,se=icar.16.sapa.1$cis$sds.rot,sort=FALSE,xlim=c(0,1
  error.dots(icar.16.sapa.2$cis$means.rot,se=icar.16.sapa.2$cis$sds.rot,sort=FALSE,xlim=c(0,1
  error.dots(icar.16.sapa.4$cis$means.rot,se=icar.16.sapa.4$cis$sds.rot,sort=FALSE,xlim=c(0,1
```

More R

R code

```
op <- par(mfrow=c(1,1)
icar.16.sapa.05$cis$mean.pair   #1238.375
icar.16.sapa.1$cis$mean.pair    # 2478.613
icar.16.sapa.2$cis$mean.pair  #  4950.1
icar.16.sapa.4$cis$mean.pair  # 9906.598

#now, just take out a general factor
 icar.16.sapa.1.05 <-fa.sapa(sapa[icar.16],n.iter=100,frac=.05)
  icar.16.sapa.1.1 <-fa.sapa(sapa[icar.16],n.iter=100,frac=.1)
  icar.16.sapa.1.2  <-fa.sapa(sapa[icar.16],n.iter=100,frac=.2)
   icar.16.sapa.1.4  <-fa.sapa(sapa[icar.16],n.iter=100,frac=.4)

error.dots(icar.16.sapa.1.05,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor   5% sample",x
error.dots(icar.16.sapa.1.1,head=40,tail=40,sort=FALSE,main="ICAR 16 1Factor 10% sample" ,xl
error.dots(icar.16.sapa.1.2,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor 20% sample",xl
 error.dots(icar.16.sapa.1.4,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor 40% sample",x


error.dots(icar.16.sapa.1.05,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor   5% ",xlim=c
error.dots(icar.16.sapa.1.1,head=40,tail=40,sort=FALSE,main="ICAR 16 1Factor 10% " ,xlim=c(0
error.dots(icar.16.sapa.1.2,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor 20% ",xlim=c(0
 error.dots(icar.16.sapa.1.4,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor 40% ",xlim=c(

icar.keys <- keys.list[398:402]


R3Diq.sapa  <-fa.sapa(sapa[icar.keys[[4]]],1,n.iter=100,frac=.05)
R3Diq.sapa.1  <-fa.sapa(sapa[icar.keys[[4]]],1,n.iter=100,frac=.1)
R3Diq.sapa.2  <-fa.sapa(sapa[icar.keys[[4]]],1,n.iter=100,frac=.2)
R3Diq.sapa.4  <-fa.sapa(sapa[icar.keys[[4]]],1,n.iter=100,frac=.4)

R3Diq.sapa$cis$mean.pair
```

R code

```
 #this next one fails
#R3Diq.sapa.01  <-fa.sapa(sapa[icar.keys[[4]]],1,n.iter=100,frac=.01)
# error.dots(R3Diq.sapa..01,head=40,tail=40,sort=FALSE,main="3D rotation items  -- 1% sample"
  op <- par(mfrow=c(2,2))
 error.dots(R3Diq.sapa,head=40,tail=40,sort=FALSE,main="3D rotation items  pairwise = 200",xl
 error.dots(R3Diq.sapa.1,head=40,tail=40,sort=FALSE,main="3D rotation items  pairwise = 399",
 error.dots(R3Diq.sapa.2,head=40,tail=40,sort=FALSE,main="3D rotation items   pairwise = 798"
  error.dots(R3Diq.sapa.4,head=40,tail=40,sort=FALSE,main="3D rotation items  pairwise=1596",

  op <- par(mfrow=c(1,1)


  samp.size <- data.frame(fraction=fraction,sample = fraction * 255348,icar16=fraction * 7434

  #now, some omega comparisons
  om16 <- omega(sapa[icar.16],4)   #the complete sample
   omega.diagram(om16)

# a 5% sample
  sapa.5.samp <- sapa[sample(1:255348,12767,replace=TRUE),icar.16]
  om.samp.5 <- omega(sapa.5.samp,4)
  omega.diagram(om.samp.5)

  mean(count.pairwise(sapa.5.samp,diagonal=FALSE),na.rm=TRUE)
  cp <- count.pairwise(sapa.5.samp)
> mean(diag(cp))
[1] 3718.562

corPlot(factor.congruence(om16,om.samp.5),numbers=TRUE,gr=gr,main="Factor Congruence: total s

sapa.02.samp <- sapa[sample(1:255348,5000,replace=TRUE),icar.16]
om.02 <- omega(sapa.02.samp,4)
 corPlot(factor.congruence(om16,om.02),numbers=TRUE,gr=gr,main="Factor Congruence: total samp
```

R code

```
#now some interesting simulations


sim.1 <- sim.irt(24,200000,low=-1,high=1,a=3)
simp <- sim.1$items
 filter <- matrix(NA,nrow=200000,ncol=24)
 filter <- sample(1:24,24*200000,replace=TRUE)

 filter<- matrix(filter,ncol=24)
simp[filter > 12 ] <- NA.  #rhis is a 50% sample
simp <- matrix(simp,ncol=24)

sim.fa.008 <- fa.sapa(simp,frac=.008,n.iter=100)
sim.fa.0016<- fa.sapa(simp,frac=.0016,n.iter=100)
 sim.fa.016<- fa.sapa(simp,frac=.016,n.iter=100)
sim.fa.032<- fa.sapa(simp,frac=.032,n.iter=100).
sim.fa.032$cis$mean.pair    #[1] 1600.788

#Now, do this again for a 25% sample
simp.25 <- sim.1$items
simp.25[filter > 6 ] <- NA.  #rhis is a 50% sample
simp,25 <- matrix(simp.25,ncol=24)
```

Error dots for factor loadings

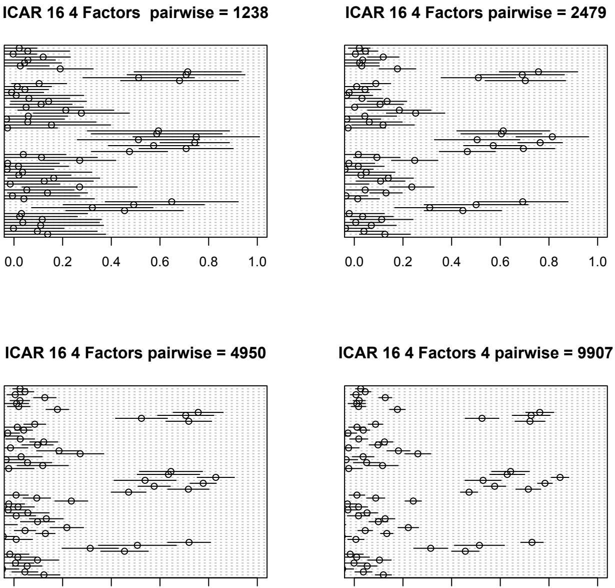
NOte that the error bars are smaller for the 25 versus 50 % samples

R code

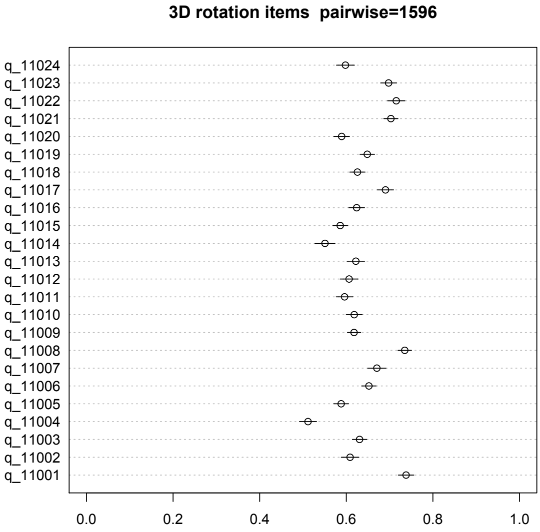
```
op <- par(mfrow=c(2,2))
error.dots(sim.fa.004,sort=FALSE,xlim=c(0,1),main="200 pairwise, 50
error.dots(sim.fa.008,sort=FALSE,xlim=c(0,1),main="400 pairwise, 50
error.dots(sim.fa.016,sort=FALSE,xlim=c(0,1),main="800 pairwise, 50
error.dots(sim.fa.032,sort=FALSE,xlim=c(0,1),main="1600 pairwise, 5

error.dots(sim.fa.25.016,sort=FALSE,xlim=c(0,1),main="200 pairwise,
error.dots(sim.fa.25.032,sort=FALSE,xlim=c(0,1),main="400 pairwise,
error.dots(sim.fa.25.064,sort=FALSE,xlim=c(0,1),main="800 pairwise,
error.dots(sim.fa.25.128,sort=FALSE,xlim=c(0,1),main="1600 pairwise
op <- par(mfrow=c(1,1))
```

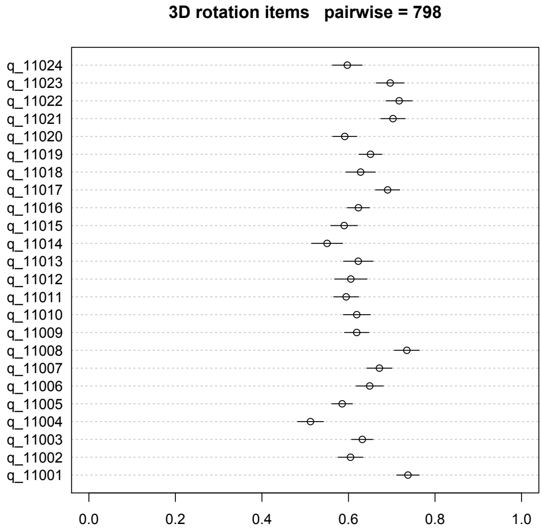
Stability of factor loadings for a 4 factor solution – Summary



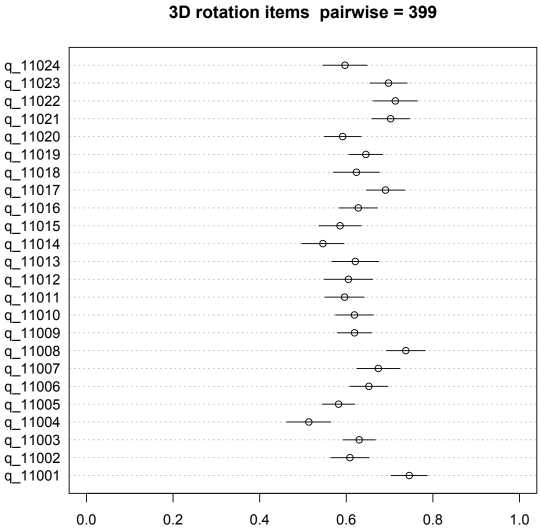
Factor loadings for the 24 3D rotation items: 40 % sample



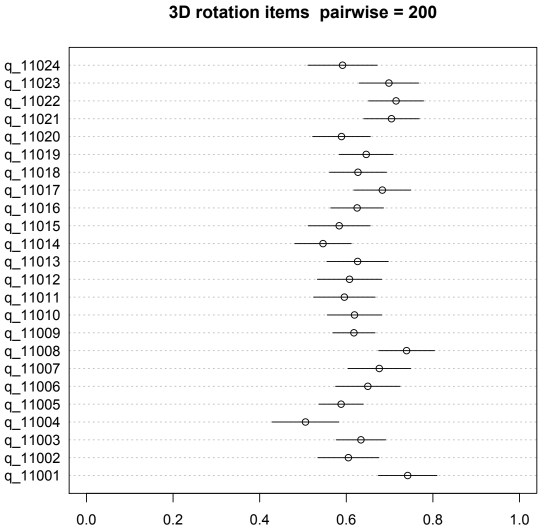
Factor loadings for the 24 3D rotation items: 20 % sample



Factor loadings for the 24 3D rotation items: 10 % sample



Factor loadings for the 24 3D rotation items: 5 % sample



Anderson, J., Lin, H., Treagust, D., Ross, S., & Yore, L. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education*, *5*(4), 591–614.

Brown, A. D. (2014). Simulating the MMCAR method: An examination of precision and bias in synthetic correlations when data are 'massively missing completely at random'. Master's thesis, Northwestern University, Evanston, Illinois.

Condon, D. . (2017). The sapa personality inventory: An empirically-derived, hierarchically-organized self-report personality assessment model.

Condon, D. M. (2014). *An organizational framework for the psychological individual differences: Integrating the affective, cognitive, and conative domains*. PhD thesis, Northwestern University.

Condon, D. M. & Revelle, W. (2014). The International Cognitive

Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52–64.

Condon, D. M. & Revelle, W. (2017). The times they are a-changin' (in personality assessment). In *European Conference on Psychological Assessment*.

Costa, P. T. & McCrae, R. R. (1992). *NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *Journal of Personality and Social Psychology*, *93*(5), 880–896.

Elleman, L., Condon, D., & Revelle, W. (2017). Exploring the ORAIS: Unexpected patterns of habits and hobbies. In *Presented as a poster*, San Antonio, Texas. Society of Personality and Social Psychology.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, &

F. Ostendorf (Eds.), *Personality psychology in Europe*, volume 7 (pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R. (2010). Personality, demographics and self reported acts: the development of avocational interest scales from estimates of the mount time spent in interest related activities. In C. Agnew, D. Carlston, W. Graziano, & J. Kelly (Eds.), *Then a miracle occurs: Focusing on the behavior in social psychological theory and research* (pp. 205–226). New York, NY: Oxford University Press.

Goldberg, L. R. & Saucier, G. (2016). The Eugene-Springfield Community Sample:
Information available from the research participants. Technical Report 56-1, Oregon Research Institute, Eugene, Oregon.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, *37*(6), 504 – 528.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big*

Five Inventory-Versions 4a and 54. Berkeley, CA: University of California, Berkeley, Institue of Personality and Social Research.

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, *70*(6), 543.

Lord, F. M. (1955). Estimating test reliability. *Educational and Psychological Measurement*, *15*, 325–336.

Lord, F. M. (1977). Some item analysis and test theory for a system of computer-assisted test construction for individualized instruction. *Applied Psychological Measurement*, *1*(3), 447–455.

Pozzebon, J. A., Visser, B. A., Ashton, M. C., Lee, K., & Goldberg, L. R. (2010). Psychometric characteristics of a public-domain self-report measure of vocational interests: The Oregon Vocational Interest Scales. *Journal of Personality Assessment*, *92*(2), 168–174.

Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, *41*(1), 203 – 212.

Revelle, W. & Condon, D. M. (2016). Embrace your missingness. In *Part of symposium: approaching complex research designs from the perspective of missing data.*, Chicago, Il. Association for Psychological Science.

Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A., & Elleman, L. G. (2016). Web and phone based data collection using planned missing designs. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *SAGE Handbook of Online Research Methods* (2nd ed.). chapter 37, (pp. 578–595). Sage Publications, Inc.

Revelle, W., Wilt, J., & Rosenthal, A. (2010). Individual differences in cognition: New methods for examining the personality-cognition link. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of Individual Differences in*

Cognition: Attention, Memory and Executive Control chapter 2, (pp. 27–49). New York, N.Y.: Springer.

Soto, C. J. & John, O. P. (2017). The next big five inventory (bfi-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of personality and social psychology*, *113*(1), 117–143.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036–1040.