Personality, Ability and Interests: Real World Outcomes

Presented as part of a symposium:

Broadening the scope of personality research: The place of Personality, Ability, and Interests in Determining Real World Outcomes

William Revelle and David M. Condon

Department of Psychology Northwestern University Evanston, Illinois USA

Slides at http://personality-project.org/sapa.html
Partially supported by a grant from the National Science Foundation: SMA-1419324

Outline

The challenge of Temperament, Abilities, and Interest

SAPA methodology
Sample items as well as people
Covariance algebra

SAPA: practice

Open source software comes to the rescue

Sample Demographics

TAI within and between majors

TAI within and between Occupations

Conclusions

Intro

Integrating Abilities interest in a broad theory of personality

- 1. Since about 1950, American personality research has tended to deemphasize (indeed, ignore) cognitive and motivational aspects of personality.
 - Researchers talk about child development and clinical diagnoses as if these were unrelated to each other and to the field of personality.
 - It is thought that young children have temperament, college students personality, clinical patients have psychopatholgy, and these should be studied as separate areas of research.
- 2. European research, on the other hand, by keeping the term "Individual Differences" alive, has continued to study these important aspects of individuality.
- 3. We attempt to continue this tradition.

Breadth vs. depth of measurement

- 1. Factor structure of domains needs multiple constructs to define structure.
- 2. Each construct needs multiple items to measure 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 with 300, the IPIP big 5 with 100, the BFI with 44 items, the TIPI with 10) to include as part of other surveys.
- 6. Particularly an issue when using large (web based) surveys, there has been a tendency to develop short forms for surveys.

Many items versus many people

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

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

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

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

4. Thus, we need to increase N as well as n. But how?

Subjects are expensive, so are items

- 1. In a survey such as Amazon's Mechanical Turk (MTURK), we need to pay by the person and by the item.
- 2. Why give each person the same items? Sample items, as we sample people.
- 3. 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)
- 4. This is the essence of Synthetic Aperture Personality
 Assessment (SAPA) (Revelle, Wilt & Rosenthal, 2010; Revelle, Condon, Wilt, French,
 Brown & Elleman, 2016)

 Method
 SAPA: practice
 Demographics
 Majors
 Occupations
 Conclusion

 O●
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O
 O

3 Methods of collecting 256 subject * items data

b) 32×8 complete c) 32×32 MCAR p=.25 a) 8×32 complete 46213634521143453443645331212414 46323114 21243623166421516154432261516513 25443314 51661351155165463622224435623344 6..3..........6.1.....6.2........5.6 43315423 11141343362332215612152135614522 26314145 25353121264561433433232246526411 414356143.2.2.....3..2.....65..5. 61335154566424114612641225353516 4223615351.....324..........23........5 24634342151536242425413513435116 62421344 11554654453123111162423325516334 35234443 ...44.4.5....3..6...6..........3.. 34514166 63415154 444413423....3.6..1.4...1..5......5. 1....54.........2.4.33..6..... 13514321 66365663 12264546 ..44...1......1..42....5..1... 31466135 ..1..3.....2..3.521......6...3.142............22.......12. 32645514 66151251 .4...2........3..162...4.....4 14411441 ..4..6..3.4...1....5.33...... 62443636 ..5..3..4...4.4..5..1......4. 33316236 63325425 4 3 . . 5 . 2 64 . 4 . . 4 . 11531126 ...1.1.2...6....4......55....2.. 61155546 3 . . 2 . . 53 2 . . 2 . 3 . 31...2..43...3.13.......5. 33245361 ...2.....4..54...2.3..62.... 52241654 63212356 24414663 ...5..3.4.....3....5.241...... 63661414 63 . 1 6 . . . 5 . . 4 . . 2 . . . 5 45555223 . . 2 . 4 . . 5 52 . 4 44 . . . 2.55.....2....6.....6.....55... 14364433 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.
- 2. Find covariances based upon "pairwise complete data".
- 3. Find scales based upon basic covariance algebra.
 - Let the raw data be the matrix X with N observations converted to deviation scores.
 - Then the item variance covariance matrix is $\mathbf{C} = \mathbf{XX}'N^{-1}$
 - and scale scores, **S** are found by S = K'X.
 - K 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, \mathbf{C}_s , is

$$\mathbf{C}_s = \mathbf{K}'\mathbf{X}(\mathbf{K}'\mathbf{X})'N^{-1} = \mathbf{K}'\mathbf{X}\mathbf{X}'\mathbf{K}N^{-1} = \mathbf{K}'\mathbf{C}\mathbf{K}. \tag{1}$$

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

SAPA is not magic: We can obtain high accuracy at the structure level but accuracy is much lower at the single subject level

- 1. Reliability of composite scales is high when formed from synthetic matrices $C_s = K'CK$ because the number of items per scale/per subject is the nominal amount.
- 2. Reliability of single scores is much less because very few items measuring a single trait are given to a single subject S = K'X.
- 3. However, the precision of the estimate of subject means (\bar{x}) is high because $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{Np-1}}$ and Np is large.
- 4. SAPA technique is very powerful for research of structure, but less powerful for research based upon single subjects.
- 5. Particularly useful in web based surveys with many subjects when we are limited in the number of items we can administer and where we are trying to increase our domain validity.

How does it work?

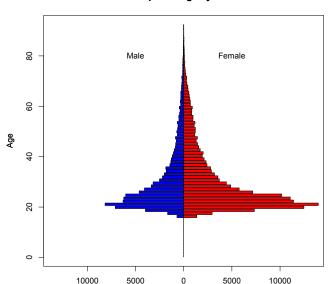
- 1. Give our basic belief in open science, we use public domain items, open source software:
 - Apache webserver, MySQL data bases, PHP and HTML5 web tools, R for statistics.
 - Extensive coding in PHP and MySQL to present item sets in random fashion (Joshua Wilt, David Condon, Jason French)
 - Code written for psychometric measurement and scale construction as implemented in the psych package (Revelle, 2016) using R (R Core Team, 2016)
- 2. Domains measured and item sources
 - Temperament items taken from International Personality Item Pool (IPIP) (Goldberg, 1999) (ipip.ori.org) and supplemented with other items.
 - Ability items have been validated (Condon & Revelle, 2014) as part of the International Cognitive Ability Resource Project (ICAR-project.org). (ICAR:Ability::IPIP:Temperament)
 - Interest items taken from Oregon Vocational Interest Survey (ORVIS) (Pozzebon, Visser, Ashton, Lee & Goldberg, 2010)

SAPA demographics

- SAPA has been running for ≈ 10 years as either personality-project.org or now sapa-project.org.
- 2. We are reporting today on the last 6 years of data based upon 229,731 non-duplicated subjects.
- In a poster present here by Lorien Elleman, we show that our results replicate differences between US states reported by Rentfrow (2010)
- 4. We have previously reported IQ data collected with SAPA as part of the International Cognitive Ability Resource project and released to the public domain (Condon & Revelle, 2014).
- 5. All analyses are done in the *psych* package (Revelle, 2016) in the open source statistical system R (R Core Team, 2016).

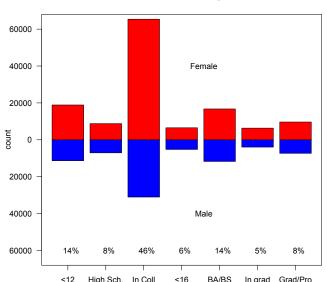
Mean age = 25.9, Median = 22, IQR = 18 - 30

Participants' Age by Gender



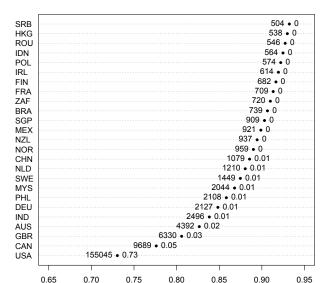
91% report their educational attainment.

Participants Education by Gender

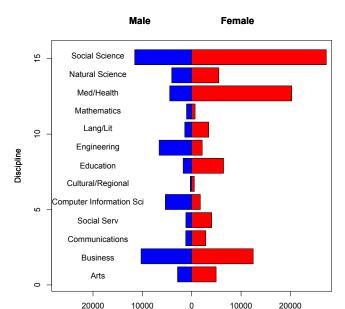


92% report country: of these 73% are US, 90% from 15 countries

25 countries account for 93% of the sample



62% report a college major. Majors by Discipline and Gender



25 majors account for 64% of those in or with college education

16.6% are in psychology

Psvchology **Business Administration and Management** Nursina Other Medicine and Allied Health Major Biology Other Social Sciences Major Accounting Health Sciences - General English Medicine (Pre-Med) Computer and Information Systems - General Elementary Education Computer Programming Marketing Political Science Finance and Financial Management Sociology Criminal Justice and Corrections Other Engineering and Technology Major Other Business Major Law and Legal Studies Social Work Health Services and Administration Other Computer and Information Sciences Major Medical Assisting



Temperament Items measures using the SAPA Personality Inventory

- 1. David Condon examined the 696 non-overlapping IPIP items that represent 18 different inventories (with 168 scales) that have what appear to be 1,894 items, In addition, those "magic 696" cover between 57% to 85% of 10 additional inventories with 235 additional scales (Condon, 2014).
- David Condon has developed a short form of 135 items that
 provides coverage of 27 different narrow domains
 (Homogeneous Item Composites) as well as five broad factors
 corresponding as much as anyone else to the traditional Big 5.
- 3. We report here analyses of Temperament, Abilities and Interests by college major and reported occupation.
- 4. All scores are found using Item Response Theory scoring of items using a quasi-Rasch model, rather than a simple sum scores of items. These two methods agree almost perfectly without missing data, but the IRT approach is more powerful with our MMCAR data.

TAI for groups is not the same as TAI for individuals

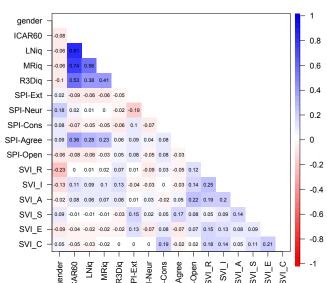
- 1. How do occupational groups or college majors differ on TAI?
 - The mean scores for groups allow us to compare the groups
 - But it is the structure of these group means that are particularly interesting for they allow us to examine niche selection based upon peoples' aptitudes and appetites.
- Overall correlation is a function of within group correlations and between group correlations.
- 3. Correlations of aggregate scores $r_{xy_{bg}}$ (between groups) \neq aggregate of correlations $r_{xy_{wg}}$ (within groups)
- 4. The overall correlation r_{xy} is a function of the within and the between correlations

$$\textit{r}_{\textit{xy}} = \textit{eta}_{\textit{x}_{\textit{wg}}} * \textit{eta}_{\textit{y}_{\textit{wg}}} * \textit{r}_{\textit{xy}_{\textit{wg}}} + \textit{eta}_{\textit{x}_{\textit{bg}}} * \textit{eta}_{\textit{y}_{\textit{bg}}} * \textit{r}_{\textit{xy}_{\textit{bg}}}$$

- These multi level correlations sometimes lead to what is known as the Yule-Simpson paradox (Kievit, Frankenhuis, Waldorp & Borsboom, 2013; Simpson, 1951; Yule, 1903)
 - These are independent and useful information.

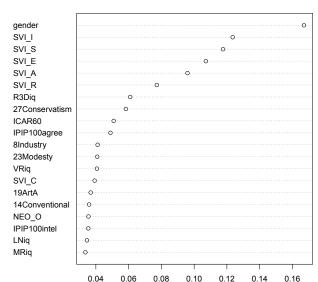
Within group correlations of Temperament, Ability, and Interests

Weighted Within Group Correlations



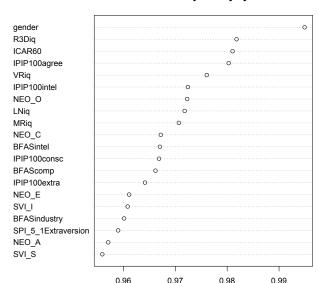
Majors: ICC1 refects the variance accounted for by group differences

ICC1 for Majors vary by TAI



Majors: ICC2 reflects the reliability of the group differences

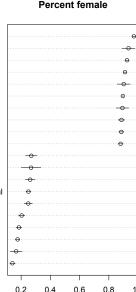
ICC2 for Majors vary by TAI



Majors differ the most in gender representation

Percent female

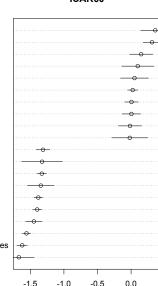
Kindergarten/Preschool Education Dance Medical Assisting **Elementary Education** Communication Disorders and Services Nursing Fashion Special Education Health Services and Administration Social Work Management Information Systems Manufacturing and Design Engineering Civil Engineering Computer and Information Systems - General **Physics** Computer Engineering Electrical Engineering Computer Programming Aerospace Engineering Mechanical Engineering



Ability difference between majors (with 95% confidence intervals)

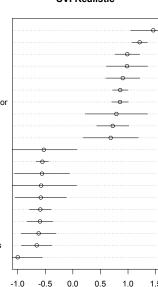
ICAR60

Statistics **Physics** Neuroscience Applied Mathematics Biomedical Engineering Computer Programming Mechanical Engineering Mathematics Industrial Engineering Materials Science and Engineering Kindergarten/Preschool Education Agricultural Businesses Social Work Culinary Arts and Sciences Criminal Justice and Corrections Health Services and Administration Physical Education Medical Assisting Human Development and Family Studies Family and Consumer Science



SVI Realistic

Manufacturing and Design Engineering Mechanical Engineering Civil Engineering Materials Science and Engineering Aerospace Engineering **Electrical Engineering** Other Engineering and Technology Major **Environmental Engineering** Biomedical Engineering Botany Sales and Marketing Operations **Elementary Education** Asian Languages and Literature Performance Studies Dance Kindergarten/Preschool Education Special Education French Communication Disorders and Services Family and Consumer Science



Artistic Interest difference between majors

0.0 0.5

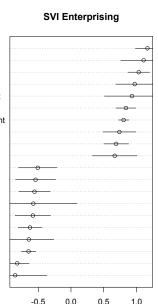
1.0

SVI Artistic

Fiction Writing Fine and Studio Arts Art History Graphic Arts Art Theory and Practice Design and Applied Arts Other Performing or Visual Art Major Drama/Theater Arts Music Comparative Literature Studies Education Administration Agricultural Businesses General Business Physical Education Hospitality Administration/Management Health Services and Administration Dentistry Medical Assisting Nursina Family and Consumer Science

Enterprising Interest difference between majors

Finance and Financial Management Entrepreneurship Marketing International Business Logistics and Supply Chain Management Accounting **Business Administration and Management** General Business **Economics** Hospitality Administration/Management Fiction Writing Art Theory and Practice Special Education Family and Consumer Science Music Education Fine and Studio Arts Communication Disorders and Services **Elementary Education** Dance French



Investigative Interest difference between majors

SVI Investigative

0.0

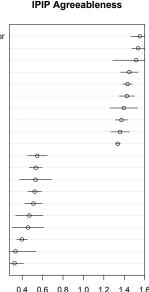
Neuroscience Chemistry Biomedical Engineering Chemical and Biological Engineering Biology Cognitive Science Botany **Physics** Materials Science and Engineering Science Education Logistics and Supply Chain Management Public Relations and Advertising Secondary Education Family and Consumer Science Other Community and Social Services Major Elementary Education Kindergarten/Preschool Education Fashion Drama/Theater Arts Sales and Marketing Operations

IPIP Agreeableness differs by major (with 95% confidence intervals)

IPIP Agreeableness

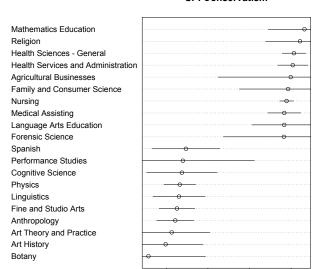
Other Community and Social Services Major Social Work Family and Consumer Science Special Education Elementary Education Human Development and Family Studies Religion Health Services and Administration Kindergarten/Preschool Education Nursing Civil Engineering Mechanical Engineering Computer Graphics Computer Engineering Mathematics Aerospace Engineering Applied Mathematics Computer Programming Asian Languages and Literature

Physics



SPI conservative differs by major (with 95% confidence intervals)

SPI Conservatism



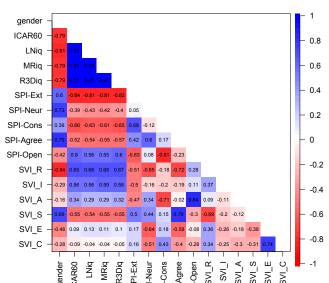
-1.0

-0.5

0.0

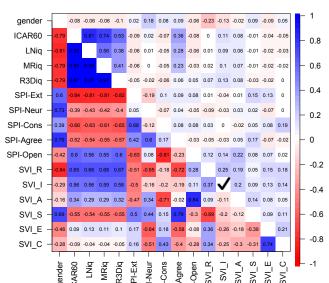
Correlation of TAI between groups is different than within groups

Weighted Between Group Correlations



Comparing TAI between groups and within groups

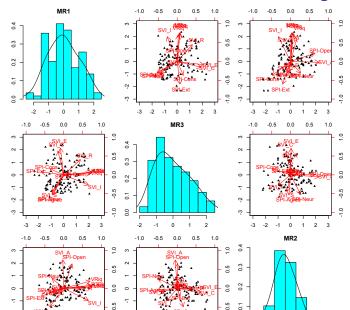
Between group and Within group correlations



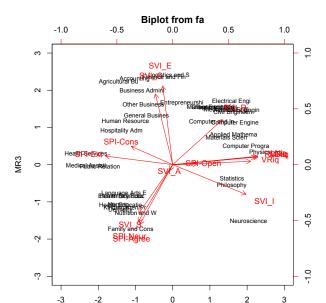
Interpreting these relationships

- 1. Students migrate into majors representing their strengths and interests
- 2. Majors choose students, (based upon ability?)
- 3. Student choose majors (based upon interests)
- 4. We can examine the factor structures of the between group correlations

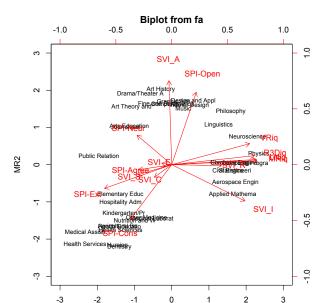
Factor structures of three dimensional between group solution



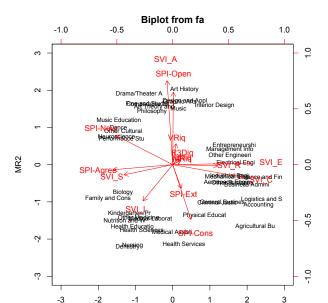
Biplot of dimensions 1 vs 2 for majors between group structure



Biplot of dimensions 1 vs. 3 for majors between group structure



Biplot of dimensions 2 vs. 3 for majors between group structure

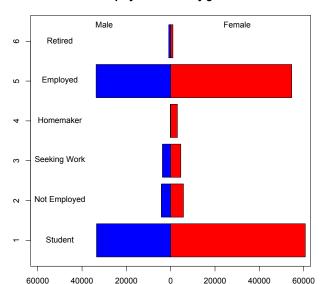


Similar results for occupations

- 1. Just as students selectively choose majors to represent their interests and abilities, so do people move into the work force to reflect their interests and abilities.
- The large group differences we see in the average personality characteristics of college majors could reflect accentuation – people become like the others in the major and small original differences accentuate into the large differences we see.

90% report their occupational status: Occupational status by Gender

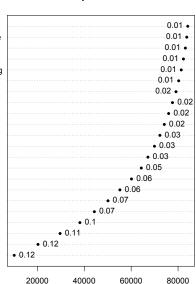
Employment status by gender



Broad Occupational "Field"

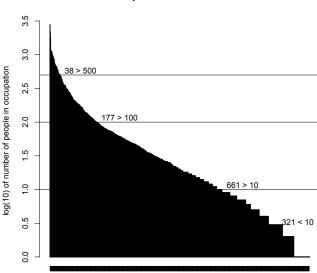
Occupational Field

FarmingFishingForestry BldqGroundsCleaningMntnce InstallMntnceRepair ConstructionExtraction TransportationMaterialMoving **ProtectiveSycs** MfgProduction LawLegalSvcs EngineeringArchitecture PersonalCareRelatedSvcs Military CommSocialSycs LifePhysicalSocialScience ArtsDesignEntnSportsMedia ComputerMath Management OfficeAdminSupport BusinFinanOperations FoodPrepServing EduTrainingLibrarySvcs Healthcare SalesRelatedSycs



Occupations are Pareto distributed with 80% in top 20%

982 Occupations are Pareto distributed



Teller

Actuary

Historian

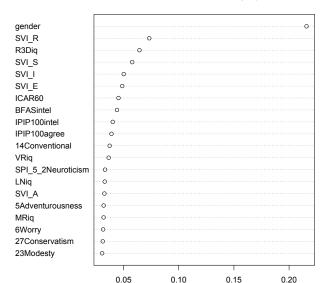
Urologist

Cashier

Actor Internist

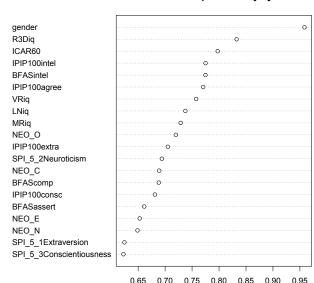
Occupations: ICC1 = variance accounted for by group differences

ICC1 for Occupation vary by TAI



Occupations: ICC2 = reliability of group differences

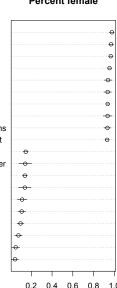
ICC2 for Occupations vary by TAI



Occupations differ by ability: top and bottom by gender

Percent female

Nanny Hairdresser, Hairstvlist, and/or Cosmetologist Preschool Teacher (except Special Education) Secretary Dental Assistant Kindergarten Teacher (except Special Education) Medical Assistant Medical Secretary Medical Records and Health Information Technicians Executive Secretary and/or Administrative Assistant Computer Programmer Other - Installation, Maintenance, and Repair Worker Computer Software Engineer Landscaping and/or Groundskeeping Worker Mechanical Engineer Network and Computer Systems Administrator Computer Systems Engineers/Architect Infantry Carpenter Automotive Mechanic and/or Service Technician

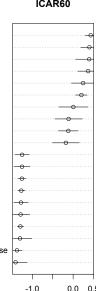


Occupations differ by ability: top and bottom by ICAR60

ICAR60

Computer Software Engineer Web Developer Software Quality Assurance Engineer and/or Tester Business Intelligence Analyst Mechanical Engineer Computer Programmer Social Science Research Assistant **Pharmacist** Instructional Designer and/or Technologist **Biologists** Other - Legal Worker Hairdresser, Hairstylist, and/or Cosmetologist Home Health Aide Medical Assistant Medical Records and Health Information Technicians Other - Legal Support Worker Nursing Aid, Orderly and/or Attendant Other - Protective Service Worker Licensed Practical Nurse and/or Licensed Vocational Nurse

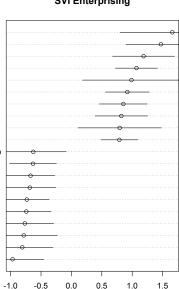
Surgical Technologist



Occupations differ by ability: top and bottom by SVI Enterprising

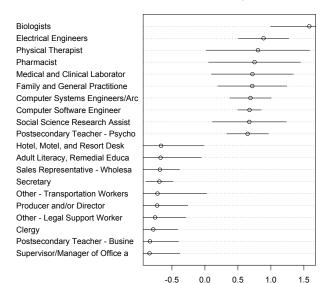
SVI Enterprising

Financial Services Sales Agent Insurance Sales Agent Financial Manager Chief Executive Officer Loan Officer Sales Manager Other - Financial Specialist Marketing Manager Sales Representative - Wholesa **Business Operations Specialist** Landscaping and/or Groundskeep Special Education Teacher - Pr Librarian Acute Care Nurse Other - Therapist Social and Human Service Assis Surgical Technologist Adult Literacy, Remedial Educa Other - Artist Medical Secretary



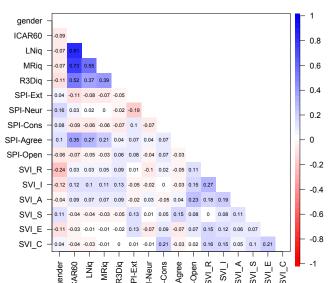
Occupations differ by ability: top and bottom by SVI Investigative

SVI Investigative



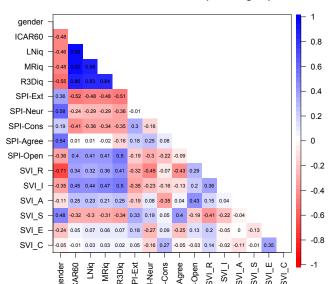
Within group correlational structure is the conventional solution

TAI correlations within Group Correlations



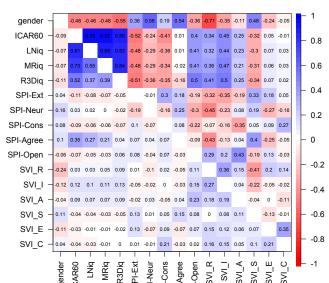
Between group correlations show a very different structure

TAI correlations between occupational groups

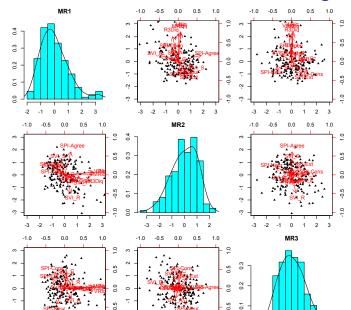


Compare the within group and between group correlations

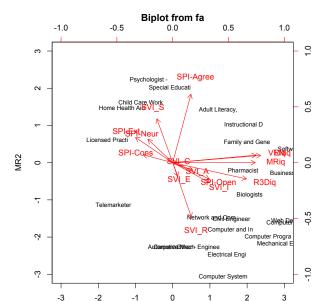
Correlation plot



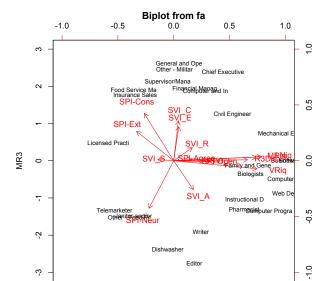
Factor structures of three dimensional between group solution



Biplot of dimensions 1 vs 2 for occupations between group structure

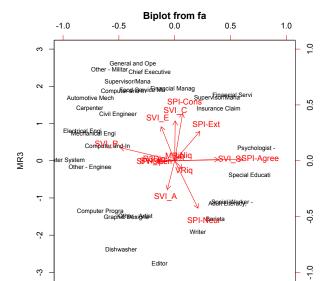


Biplot of dimensions 1 vs. 3 for occupations between group structure



53 / 57

Biplot of dimensions 2 vs. 3 for occupations between group structure



Conclusions

Expanding the Personality toolbox: Abilities and Interest

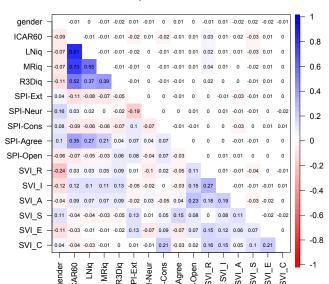
- 1. When predicting real world outcomes such as choice of college major or occupation, it is important to go beyond traditional personality measures.
- 2. Ability serves as filter to college majors and occupations
- 3. Interests direct choice between majors and occupations.
- 4. Personality, ability and temperament structures at group level are very different than those within groups.

More information

- Slides are at http://personality-project.org/sapa.html
- Ability measures are taken from the International Cognitive Ability Resource (Condon & Revelle, 2014) (see http://icar-project.com)
- 3. Data sets are available at DataVerse: (Condon & Revelle, 2015).
- Analytical code done using the psych package (Revelle, 2016) in R (R Core Team, 2016).

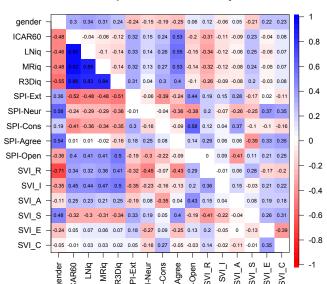
Compare the within group correlations for occupations and majors

Correlations within occupations do not differ from within n



Compare the between group correlations for occupations and majors

Between occupations and between majors do differ



The following is not included as slides but is included in the pdf to help see how to do the analysis.

First, some basic descriptives

```
R code
tedu <- table(demo.TAI$education)</pre>
totedu <- sum(!is.na(demo.TAI$education))
tedu/totedu
with (demo.TAI, bi.bars (age, as.numeric (gender), main="Participants|
text (-5000, 80, "Male")
text(5000,80,"Female")
with (demo.TAI, bi.bars (education, as.numeric (gender)))
with (demo.TAI, bi.bars (as.numeric (education), as.numeric (gender), horiz=
text(5,40000, "Female")
text(5,-40000, "Male")
text(.8,-60000,"14%")
text (2,-60000, "8%")
text (3.2, -60000, "46%")
text (4.4, -60000, "6%")
text(5.6,-60000,"14%")
text(6.8,-60000,"5%")
text(8,-60000,"8%")
```

```
sum(!is.na(demo.TAI$education))/nrow(demo.TAI)
tc <- table(demo.TAI$country)</pre>
totc <- sum(!is.na(demo.TAI$country))</pre>
tcs <- sort(tc,TRUE)
dotchart(tcs[2:35], main="Count by country (US = 155,045 not shown)")
for (i in 3:35) {text(tcs[i],i-1,tcs[i],pos=4)}
text(tcs[2],1,tcs[2],pos=2)
ctcs <- cumsum(tcs)
dotchart(ctcs[1:25]/totc,main="25 countries account for 93% of the sa
#text(ctcs[1]/tot,1,tcs[1],pos=4)
 for (i in 1:25) {text(ctcs[i]/totc,i,tcs[i],pos=2)
text(ctcs[i]/totc,i,round(tcs[i]/totc,2), pos=4)
tcs[1]/tot
 ctcs[101/tot
disc.name <- c("Arts", "Business", "Communications", "Social Serv", "Comp
with (demo.TAI, bi.bars (as.numeric (discipline), as.numeric (gender), hori
 for(i in 1:13) {
 text (-19000, (i-.5)*15.8/13, disc.name[i], srt=0)}
```

```
with (demo.TAI, bi.bars (as.numeric (jobstatus), gender, horiz=TRUE, main =
job.names <- c("Student", "Not Employed", "Seeking Work", "Homemaker", "
for(i in 1:6) {text(-50000,(i-.4) * 6/5,job.names[i],srt=00)}
text(-30000,7.2,"Male")
text(30000,7.2,"Female")
sum(!is.na(demo.TAI$jobstatus))/nrow(demo.TAI)
maj <- table(demo.TAI$major)</pre>
maj <- sort(maj, TRUE)</pre>
dotchart(mai[25:1],main=" 16.6% are in psychology"))
cmai <- cumsum(mai)
tot <- sum(maj)
pmaj<- cmaj/tot
tot/nrow(demo.TAI)
sum(mai > 100)
plot(cmaj,xlab="College Major",ylab="Cumulative number of majors")
plot(log10(maj),xlab="College Major",ylab="Log (10) of number in major"
abline(h=3)
abline(h=2)
text(118, 2.05, "118 > 100 ", pos=4)
text(38,3.05,"38 > 1,000",pos=4)
```

```
abline (h=log10 (500))
text(68,log10(500)+.05,"68 > 500",pos=4)
if <- table(demo.TAI$iobfield)</pre>
jf <- sort(jf,TRUE)</pre>
cif <- cumsum(if)
totif
dotchart(cjf,main = "Occupational Field", pch=20)
for(i in 1:15) {text(cjf[i],i, round(jf[i]/totjf,2), pos=4)}
for(i in 16:22) {text(cjf[i],i, round(jf[i]/totjf,2), pos=2)}
occ <- table(demo.TAI$occupation)</pre>
욧
 occ <- sort (occ, TRUE)
plot(log10(occ[1:200]), main="Top 200 occupations account for 80% of
 abline (h=log10(100))
 text(177, log10(100) + .05, "177 > 100", pos=4)
 abline(h=log10(500))
text(38,log10(500)+.05,"38 > 500",pos=4)
plot(log10(occ), main="982 Occupations are Pareto distributed", ylab="le
abline (h=log10(100))
```

```
text(177,log10(100)+.05, "177 > 100", pos=4)
abline(h=log10(500))
text(38,log10(500)+.05, "38 > 500",pos=4)
abline(h=log10(10))
text(651,1.05, "661 > 10", pos=4)
text(860,.5, "321 < 10",pos=4)
```

Now, lets do it by major

```
sb.demo.TAI <- statsBy(demo.TAI[c(1:30,36:93)],group="major")
names(sb.demo.TAI)
[1] "mean" "sd" "n" "F" "ICC1" "ICC2" "raw" "rbg"

icc1 <- sb.demo.TAI$ICC1
names(icc1) <- sub(".theta","",names(icc1))
names(icc1) <- sub("SPI_27_","",names(icc1))
icc1.p <- icc1[-c(1:10,12:25)]
dotchart(sort(icc1.p)[c(45:64)],main="ICC1 for Majors vary by TAI")
icc2 <- sb.demo.TAI$ICC2
```

```
names(icc2) <- sub(".theta", "", names(icc2))</pre>
 names(icc2) <- sub("SPI 27 ","",names(icc2))</pre>
 icc2.p <- icc2[-c(1:10,12:25)]
 dotchart(sort(icc2.p)[c(45:64)], main="ICC2 for Majors vary by TAI")
maj.mean <- sb.demo.TAI$mean
mai.n <- sb.demo.TAI$n
maj.sd <- sb.demo.TAI$sd
colnames(mai.mean) <- sub(".theta","",colnames(mai.mean))</pre>
colnames(maj.mean) <- sub("SPI 27 ","",colnames(maj.mean))</pre>
colnames(maj.sd) <- sub("SPI 27 ","",colnames(maj.sd))</pre>
colnames(mai.sd) <- sub(".theta", "", colnames(mai.sd))</pre>
maj.se <- maj.sd /sqrt(maj.n)</pre>
mai.mean100 \leftarrow subset(mai.mean,(mai.n[,26] > 99))
maj.n100 <- subset(maj.n, (maj.n[,26] > 99))
maj.sd100 <- subset(maj.sd, (maj.n[,26] > 99))
maj.se.100 \leftarrow subset(maj.se,(maj.n[,26] > 99))
```

 $maj.data \leftarrow list(mean=maj.mean100, n = maj.n100, sd = maj.sd100)$

dotchart.psych(maj.mean100[ord[c(1:10,108:117)], "gender"]-1, maj | se.10

ord <- order(maj.mean100[, "gender"])</pre>

ord <- order(maj.mean100[,"ICAR60"])</pre>

```
dotchart.psych(maj.mean100[ord[c(1:10,108:117)],"ICAR60"],maj.se.100
dotchart(sort(maj.mean100[,"ICAR60"])[c(1:10,108:117)], main="ICAR 60
ord <- order(maj.mean100[,"SVI I"])</pre>
dotchart.psych(maj.mean100[ord[c(1:10,108:117)], "SVI I"], maj.se.100[
 #dotchart(sort(maj.mean100[,"SVI_I"])[c(1:10,108:117)], main="$VI In:
 ord <- order(maj.mean100[,"SVI E"])</pre>
dotchart.psych(maj.mean100[ord[c(1:10,108:117)], "SVI_E"], maj.se.100[
 #dotchart(sort(maj.mean100[, "SVI E"])[c(1:10,108:117)], main="SVI En
ord <- order(maj.mean100[,"SVI_A"])</pre>
dotchart.psych(maj.mean100[ord[c(1:10,108:117)], "SVI_A"], maj.se.100[
# dotchart(sort(mai.mean100[,"SVI A"])[c(1:10,108:117)], main=|SVI A
ord <- order(maj.mean100[,"SVI_R"])</pre>
dotchart.psych(maj.mean100[ord[c(1:10,108:117)], "SVI_R"], maj.se. 100[o
# dotchart(sort(maj.mean100[, "SVI_R"])[c(1:10,108:117)], main="$VI Re.
ord <- order(maj.mean100[,"27Conservatism"])</pre>
                                                                   57 / 57
```

```
dotchart.psych(maj.mean100[ord[c(1:10,108:117)],"27Conservatism"],maj
dotchart(sort(maj.mean100[, "27Conservatism"])[c(1:10,108:117)], | main=
ord <- order(mai.mean100[,"IPIP100agree"])</pre>
dotchart.psych(maj.mean100[ord[c(1:10,108:117)],"IPIP100agree"] \mid maj.s
dotchart(sort(maj.mean100[,"IPIP100agree"])[c(1:10,108:117)], main="I
rbg.maj <- sb.demo.TAI$rbg
 colnames(rbq.maj) <- rownames(rbq.maj) <- sub(".theta.bq", "", colname
  colnames(rbg.maj) <- rownames(rbg.maj) <- sub("SPI_27_","",colname
  colnames(rbq.maj) <- rownames(rbq.maj) <- sub("SPI_5_","",colnames</pre>
    colnames(rbq.maj) <- rownames(rbq.maj) <- sub(".bq","",colnames(</pre>
  colnames(rbg.maj)[57:61] <- c("SPI-Ext", "SPI-Neur", "SPI-Cons", "SPI-
  colnames(rbq.maj) <- rownames(rbq.maj) <- sub("IPIP100", "IPIP", col
   rownames(rbg.maj) <- colnames(rbg.maj)</pre>
 corPlot(rbg.maj[c(10,25:28,57:61,82:87),c(10,25:28,57:61,82:87)],upp
 rwg.mai<- sb.demo.TAI$rwg
 colnames(rwg.maj) <- rownames(rwg.maj) <- sub(".theta.wg", "", colna
```

colnames(rwg.maj) <- rownames(rwg.maj) <- sub("SPI_27_","",colname
colnames(rwg.maj) <- rownames(rwg) <- sub("SPI_5_","",colnames(rwg
colnames(rwg.maj) <- rownames(rwg.maj) <- sub(".wg","",colnames(</pre>

```
colnames(rwg.maj)[57:61] <- c("SPI-Ext", "SPI-Neur", "SPI-Cons", "SP
    rownames(rwg.maj) <- colnames(rwg.maj)
corPlot(rwg.maj[c(10,25:28,57:61,82:87),c(10,25:28,57:61,82:87)], upp</pre>
```

What about the factor structure of TAI between groups.

biplot (f3.bg, labels=maj.names, cuts=1.5)

```
R code

f3.bg<- fa(rbg[c(26:29,57:61,82:87),c(26:29,57:61,82:87)],3)

ov <- c(26:29,57:61,82:87)
f3.bg <- fa(rbg.maj[ov,ov],3)
maj.f3.scores <- factor.scores(maj.mean100[,ov+1],f3.bg) #because the maj.names <- rownames(f3.bg$scores$scores)
maj.names <- substr(maj.names,1,15)

f3.bg$scores <- maj.f3.scores
biplot(f3.bg,labels=maj.names,choose=c(1,2),cuts=1.5)
```

occ.sd <- sb.demo.TAI.occ\$sd

Now, for occupation: Basically just a repeat of the major procedures

```
R code
sb.demo.TAI.occ <- statsBy(demo.TAI[c(1:30,36:93)],group="occupation"
icc1 <- sb.demo.TAI.occ$ICC1
names(icc1) <- sub(".theta", "", names(icc1))</pre>
 names(icc1) <- sub("SPI 27 ", "", names(icc1))
 icc1.p <- icc1[-c(1:10,12:25)]
 dotchart (sort (iccl.p) [c(45:64)], main="ICCl for Occupation vary by TA
  icc2 <- sb.demo.TAI.occ$ICC2
 names(icc2) <- sub(".theta", "", names(icc2))</pre>
 names(icc2) <- sub("SPI 27 ","",names(icc2))</pre>
 icc2.p <- icc2[-c(1:10,12:25)]
 dotchart (sort (icc2.p) [c(45:64)], main="ICC2 for Occupations vary by T.
 occ.mean <- sb.demo.TAI.occ$mean
occ.n <- sb.demo.TAI.occ$n
```

```
colnames(occ.mean) <- sub(".theta", "", colnames(occ.mean))</pre>
colnames(occ.mean) <- sub("SPI 27 ","",colnames(occ.mean))</pre>
colnames(occ.sd) <- sub("SPI 27 ","",colnames(occ.sd))</pre>
colnames(occ.sd) <- sub(".theta","",colnames(occ.sd))</pre>
occ.se <- occ.sd /sqrt(occ.n)
occ.mean100 \leftarrow subset(occ.mean,(occ.n[,26] > 99))
occ.n100 \leftarrow subset(occ.n,(occ.n[,26] > 99))
occ.sd100 \leftarrow subset(occ.sd,(occ.n[,26] > 99))
occ.se.100 \leftarrow subset(occ.se,(occ.n[,26] > 99))
occ.names <- rownames(occ.mean100)
occ.names <- substr(occ.names, 1, 30)
rownames(occ.mean100) <- occ.names
occ.data <- list(mean=occ.mean100,n = occ.n100,sd = occ.sd100)
n.occ <- 176
ord <- order(occ.mean100[, "gender"])</pre>
dotchart.psych(occ.mean100[ord[c(1:10,(n.occ-9):n.occ)],"gender']-1,o
ord <- order(occ.mean100[,"ICAR60"])</pre>
dotchart.psych(occ.mean100[ord[c(1:10,(n.occ-9):n.occ)],"ICAR60|,],occ.
```

```
ord <- order(occ.mean100[,"ICAR60"],decreasing=TRUE)
dotchart.psych(occ.mean100[ord[20:1],"ICAR60"],occ.se.100[ord[1:20],"
dotchart(sort(occ.mean100[,"ICAR60"])[c(1:10,(n.occ-9):n.occ)], | main=
ord <- order(occ.mean100[,"SVI_I"])</pre>
dotchart.psych(occ.mean100[ord[c(1:10,(n.occ-9):n.occ)],"SVI_I"|],occ
#dotchart(sort(occ.mean100[, "SVI I"])[c(1:10, (n.occ-9):n.occ)] | main
ord <- order(occ.mean100[,"SVI E"])</pre>
dotchart.psych(occ.mean100[ord[c(1:10,(n.occ-9):n.occ)], "SVI E"], occ
rbq.occ <- sb.demo.TAI.occ$rbq
colnames(rbg.occ) <- rownames(rbg.occ) <- sub(".theta.bg", "", colname
  colnames(rbq.occ) <- rownames(rbq.occ) <- sub("SPI_27_","",colname
  colnames(rbq.occ) <- rownames(rbq.occ) <- sub("SPI_5_","",colnames
    colnames(rbq.occ) <- rownames(rbq.occ) <- sub(".bq", "", colnames(
  colnames(rbq.occ)[57:61] <- c("SPI-Ext", "SPI-Neur", "SPI-Cons", "SPI-
  colnames(rbg.occ) <- rownames(rbg.occ) <- sub("IPIP100", "IPIP", col
   rownames(rbg.occ) <- colnames(rbg.occ)
  rwg.occ <- sb.demo.TAI.occ$rwg
colnames(rwq.occ) <- rownames(rwq.occ) <- sub(".theta.bq","",colname
  colnames(rwq.occ) <- rownames(rwq.occ) <- sub("SPI_27_","",colname
  colnames(rwg.occ) <- rownames(rwg.occ) <- sub("SPI_5_","",colnames
```

colnames(rwg.occ) <- rownames(rwg.occ) <- sub(".bg","",colnames(colnames(rwg.occ)[57:61] <- c("SPI-Ext", "SPI-Neur", "SPI-Cons", "SPI-colnames(rwg.occ) <- rownames(rwg.occ) <- sub("IPIP100", "IPIP", colnames(rwg.occ) <- sub("IPIP100", "IPIP100", "IPIP10

```
rownames(rwg.occ) <- colnames(rwg.occ)
corPlot(rbg.occ[c(10,25:28,57:61,82:87),c(10,25:28,57:61,82:87)],upper
corPlot(rwg.occ[c(10,25:28,57:61,82:87),c(10,25:28,57:61,82:87)],upper
occ.upperlower <- lowerUpper(rwg.occ[c(10,25:28,57:61,82:87),c(10,25:28,57:61,82:87),c(10,25:28)]
corPlot(occ.upperlower,numbers=TRUE)
```

Now, a few comparisons of grouping by major and grouping by occupatio

```
occ.maj.upperlower <- lowerUpper(rwg.occ[c(10,25:28,57:61,82:87),c(10corPlot

occ.maj.upperlower <- lowerUpper(rwg.occ[c(10,25:28,57:61,82:87),c(10corPlot(occ.maj.upperlower,numbers=TRUE,main="Within occupations and occ.maj.upperlower.rbg<- lowerUpper(rbg.occ[c(10,25:28,57:61,82:87),c(10corPlot(occ.maj.upperlower.rbg,numbers=TRUE,main="Between occupations"), corPlot(occ.maj.upperlower.rbg,numbers=TRUE,main="Between occupations")
```

What about the factor structure of TAI between occupational groups.

```
R code
occ.f3.bg<- fa(rbg.occ[c(26:29,57:61,82:87),c(26:29,57:61,82:87)],3)
ov \leftarrow c(26:29,57:61,82:87)
occ.f3.bg <- fa(rbg.occ[ov,ov],3)
occ.f3.scores <- factor.scores(occ.mean100[.ov+1].occ.f3.bg)
                                                                #because
occ.names <- rownames(occ.f3.bg$scores$scores)
occ.names <- substr(occ.names, 1, 15)
occ.f3.bg$scores <- occ.f3.scores
biplot(occ.f3.bg,labels=occ.names,choose=c(1,2),cuts=1.8)
biplot(occ.f3.bg,labels=occ.names,cuts=1.8)
biplot (occ.f3.bg,labels=occ.names,choose=c(1,3),cuts=1.8)
biplot(occ.f3.bg,labels=occ.names,choose=c(2,3),cuts=1.8)
```

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. (2015). Selected personality data from the SAPA-Project: 08dec2013 to 26jul2014. *Harvard Dataverse*.
- 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.
- Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*, 4(513), 1–14.
- 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.
- R Core Team (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Rentfrow, P. J. (2010). Statewide differences in personality: Toward a psychological geography of the united states. *American Psychologist*, *65*(6), 548–558.
- Revelle, W. (2016). psych: Procedures for Personality and Psychological Research. http://cran.r-project.org/web/packages/psych/: Northwestern University, Evanston. R package version 1.6.4.
- 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* chapter 37: Mobile Methods. Sage Publications, Inc. (in press).

- 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.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society.* Series B (Methodological), 13(2), 238–241.
- Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika*, 2(2), 121–134.