Open science <=> open methods + open data

Presented (virtually) at the
4th International Conference on
Artificial Intelligence and Big Data in Education
Beijing, China

William Revelle
Northwestern University
Evanston, Illinois USA

Slides and supplementary material available at
personality-project.org/sapa

December 12, 2020

Outline

Abstract

There are two complementary approaches to the open science of ability measurement:

1. The development of the International Cognitive Ability Resource (ICAR), an international collaboration to develop, validate and distribute an open source measure of cognitive ability.
2. The second development which has facilitated international research in ability is the use of the open source statistical system, R and the more than 16,000 packages developed with R.

Using ICAR items from an open source database as an example, I will show how various R packages can be used for classical and IRT based assessment of ability.

Open Science: A new idea or a long term tradition?

1. Science is a process for asking questions that have answers
   - Our questions and our answers need to be open and shared.
   - Our way of addressing these questions should be open to all.
   - Our results are for everyone, not just those who can afford to pay for journals.
   - Our results need to trusted and trustworthy.
2. This is not a new idea, sharing ideas, methods and results is as old as the British Royal Society from 1660.
   - The Royal Society was an 'invisible college' of natural philosophers and physicians.
   - Royal Society's motto 'Nullius in verba' is taken to mean 'take nobody's word for it'. (We might now say, does it replicate?)
3. The study of cognitive ability can and should become an example of open science.
   - Traditional studies have been been well powered and replicable, but unfortunately, have been propietary.
   - However, there is now a growing tendency to use open and shared materials.

Open science involves materials, methods, and publications

Open science means sharing our questions, our methods, our materials, and our publications. I address these issues with two examples:

1. Open materials for the assessment of cognitive ability: the International Cognitive Ability Resource (ICAR) project.
2. Open software for statistical analysis of data: particularly the *psych* package for the the R project.

Individuals differ in cognitive ability

1. Ever since antiquity, people have used measures of cognitive ability for selection and prediction.
2. Plato stated that that leaders should show exceptional ability.
3. Theophrastus depicted those with low ability ("the stupid man") as slow in speech and action.
4. Confucius classified people on the basis of intelligence as to whether they were of great wisdom, average intelligence, or of little intelligence.
5. Mencius's saying that "scaling makes it possible to understand weight, measurement makes it possible to understand length; these are true for all things, especially true for the mind" may be seen as "a clear and important statement of the importance of quantitative measurement of the human mind"(p 102, Zhang, 1988).

Cognitive Testing: the first 2000 years

1. Most American scholars attribute the first measurement of cognitive ability to Binet and Simon (Binet & Simon, 1905) and subsequently to Louis Terman (Terman, 1916).
2. However, earlier work in the UK was that of Francis Galton (Galton, 1869) (see Jensen, 2002, for details).
3. Some of the most influential work was done in the UK by Charles Spearman who developed not just factor analysis but also a general theory of intelligence (Spearman, 1904).
4. Much earlier use of cognitive assessment for personnel selection was found starting with the Han Dynasty of China in approximately 200 BCE (Bowman, 1989; Urbina, 2014; Zhang, 1988).
5. Some of us attribute an even earlier use of sequential assessment to Gideon, whose story is told in the Hebrew Bible (Judges 7) as well as to Plato in *Republic* (VII: 534, 537).

Cognitive testing the 20th century

1. Until 1915, with the exception of the Chinese assessments for the Civil Service (Zhang, 1988) assessments of ability required one on one assessments.
2. In response to the need to select millions of men for the U.S. Army, Yerkes and his colleagues (Yoakum & Yerkes, 1920) developed paper and pencil forms (the *Army Alpha*) to screen potential recruits for training.
3. These procedures for military assessment were expanded for recruit classification in the Second World War (Dubois, 1947; Murray, MacKinnon, Miller, Fiske & Hanfmann, 1948).
4. Other large scale, proprietary measures were developed for college (the SAT and ACT) and graduate (GRE) and medical school admission (MCAT) (Kuncel & Hezlett, 2007).

Validity and prediction studies

1. That these tests work has been well established (Deary, 2012; Haier, 2016; Hunt, 2010; Sackett, Lievens, Van Iddekinge & Kuncel, 2017; Sackett & Kuncel, 2018).
2. Measures taken in childhood and early adolescence are stable over at least 7 decades, and predict life expectancy (Deary, 2008), success in school, job performance, marital stability, and social mobility (Gottfredson, 1997).
3. Ability tests taken in high school predict occupational status 50 years later (Damian, Spengler, Sutu & Roberts, 2019; Spengler, Damian & Roberts, 2018).
4. Highly able 10-12 year olds have very successful life time careers (Terman & Oden, 1947, 1959).
5. Adolescents whose scores on the SAT are equivalent to those several years older go onto stellar careers in the humanities and sciences (Lubinski & Benbow, 2006; Lubinski, Benbow, Webb & Bleske-Rechek, 2006).

Unfortunately, all of these tests have been proprietary

1. From the very beginning of modern testing, the tests were developed by government or large industries and were not open to broad use.
2. For security reasons, test developers kept the items confidential.
3. For profit reasons, tests were limited to qualified administrators.
4. Tests were given as paper and pencil to groups or individually.
5. Researchers had to pay to get access to these tests.
6. Even the quasi open ETS "French Kit" charged $0.15 per copy for graduate students which limited sample sizes.
7. Studying cognitive ability was thus limited to a select few with adequate funding and resources.
8. The web has changed all of this (Revelle, Dworak & Condon, 2020).

The International Cognitive Ability Resource (ICAR)

1. Combining the items from an honors thesis project (Liebert, 2006) with prior work on geometric analogies (Leon & Revelle, 1985) and the development of 3D rotation items, the International Cognitive Ability Resource (ICAR) was born (Condon, 2012; Condon & Revelle, 2014).
2. The original test development was part of the personality-project.org and was an early application of our Massively Missing Completely at Random (MMCAR) design (Revelle, Wilt & Rosenthal, 2010; Revelle, Condon, Wilt, French, Brown & Elleman, 2016).
3. Items were designed to be “Google resistant” (answers could not be looked up) and to be self administered over the web (i.e., including cell phones).
4. Further improvements to the web design and item set followed and the test moved to sapa-project.org (Condon, 2018).
5. Following a public release of the original pool of 60 items (Condon & Revelle, 2014), an international consortium was formed to further develop the item set.

A brief history of the ICAR

1. Original test development had 60 items spanning four constructs (Condon, 2012):
    - Verbal reasoning
    - Letter and number series
    - Matrix reasoning
    - Three dimensional rotations
2. As described by (Condon, 2012) this test was validated against self reported SAT/ACT scores (N =34,229) and the 16 item sample test against the Shipley-2 measure of cognitive functioning (Shipley, 2009) (N=137).
3. Sample items are given in the original article (Condon & Revelle, 2014). Please do not reproduce these with the answers!
4. I show examples of the four item types, and their factor structure (Revelle, 2020a)
5. For a more detailed history of the development (Condon, 2012; Condon & Revelle, 2014) and use of the ICAR, see (Revelle et al., 2020).

Sample ICAR Verbal Reasoning

VR.4
What number is one fifth of one fourth of one ninth of 900?
(1) 2 (2) 3 (3) 4 (4) 5 (5) 6 (6) 7

VR.16
Zach is taller than Matt and Richard is shorter than Zach. Which of the following statements would be most accurate?
(1) Richard is taller than Matt (2) Richard is shorter than Matt (3) Richard is as tall as Matt (4) It's impossible to tell

---

VR.17
Joshua is 12 years old and his sister is three times as old as he. When Joshua is 23 years old, how old will his sister be?
(1) 35 (2) 39 (3) 44 (4) 47 (5) 53 (6) 57

VR.19
If the day after tomorrow is two days before Thursday then what day is it today?
(1) Friday (2) Monday (3) Wednesday (4) Saturday (5) Tuesday (6) Sunday

---

Sample ICAR Letter Number Reasoning

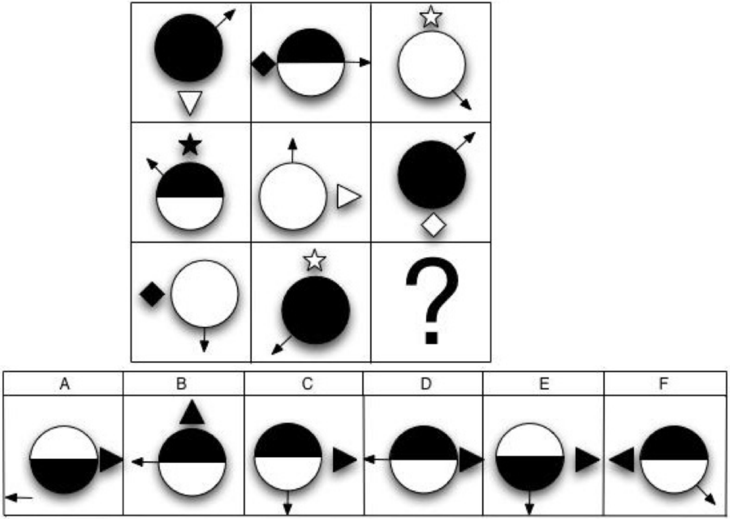
LN.7
In the following alphanumeric series, what letter comes next? K N P S U
(1) S (2) T (3) U (4) V (5) W (6) X

LN.33
In the following alphanumeric series, what letter comes next? V Q M J H
(1) E (2) F (3) G (4) H (5) I (6) J

LN.34
In the following alphanumeric series, what letter comes next? I J L O S
(1) T (2) U (3) V (4) X (5) Y (6) Z

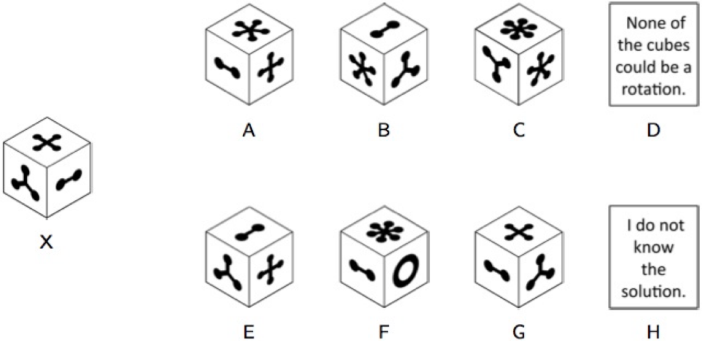
LN.58
In the following alphanumeric series, what letter comes next? Q S N P L
(1) J (2) H (3) I (4) N (5) M (6) L

Sample ICAR Spatial Analogies



Sample ICAR 3 Dimensional Rotation ability

R3D.3 All the cubes below have a different image on each side. Select the choice that could represent a rotation of the cube labeled X.

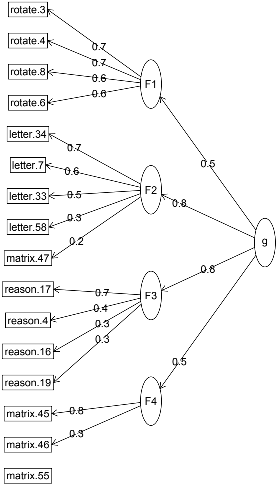


Sample data sets are publicly available

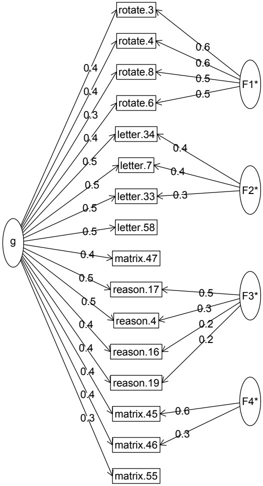
1. 4,000 cases for the 16 items from the ICAR sample test are included in the *psychTools* package (Revelle, 2020b).
2. 96,958 cases for the 60 items were discussed in (Condon & Revelle, 2014) and are available for download from the *Journal of Open Psychology Data* (Condon & Revelle, 2016).
3. Factor structures of these items show a fairly clean four correlated factors with a higher order 'g'.

Hierarchical and SL transformations of the ICAR sample test

Hierarchical (multilevel) Structure



Schmid Leiman Transformed



The International Cognitive Ability Resource project

1. With funding from 3 different counties coordinated by the ORA plus in Europe, the ICAR project was born and received funding for three years. It continues as an open science consortium.
2. Principal investigators included
   - Philipp Doebler from TU Dortmund University, and Heinz Holing from the University of Munster (funded by the DFG)
   - John Rust, David Stillwell, and Luning Sun from Cambridge University in the UK. (Funded by the ESRC)
   - William Revelle and David Condon from Northwestern University and the University of Oregon in the US (NSF)
3. International collaborators included Professor Fang Luo from Beijing Normal University and Ricardo Primi from the University of Sao Francisco, Brazil.
4. Other investigators are welcome to join us at the International Cognitive Ability Resource (ICAR) (Condon, Doebler, Holling, Gühne, Rust, Stillwell, Sun, Chan, Loe & Revelle, 2014)

Further developments available on ICAR

Members and associates of the ICAR team have been developing a total of 19 item types and have administered over 1,000 different items.

1. An automated perceptual maze test (Loe & Rust, 2017)
2. Automated number series item generator (Loe, Sun, Simonfy & Doebler, 2018)
3. Online Spatial Network Measures (Loe, 2020)
4. Dynamic Propositional Reasoning (Gühne, Doebler, Condon, Luo & Sun, 2020)
5. A sample of Item types still being validated
    - Automatic Compound Remote Associates
    - Two-Dimensional Rotations
    - Figural Analogies
    - Emotion recognition
    - Arithmetic
    - Face-detection (aka the the Mooney Test)
    - A situational judgement task.

Many different researchers have started using the ICAR items and tests

1. As of January, 2020, 79 publications from many different investigators had administered ICAR items (Dworak, Revelle, Doebler & Condon, 2020).
2. This number has increased over the past year with investigators sometimes using just a few ICAR items (Sobkow, Olszewska & Traczyk, 2020), and sometmes selecting larger subsets (Svenson, Guillen & others, 2020).
3. An independent study was a further validation and reported the correlation of the total score of the 16 item sample test with the full score from the WAIS was .81 with a latent factor correlation of .94 (Young & Keith, 2020).
4. The most important point of all of these studies is that the ICAR items are freely available and can be self administered over the web.

A few representative studies

1. MTurk samples do not seem to differ from the original web based derivation sample (Merz, Lace & Eisenstein, 2020).
2. ICAR used in a national panel study ("How nuts are the Dutch") (Krieke, Jeronimus, Blaauw, Wanders, Emerencia, Schenk, Vos, Snippe, Wichers, Wigman & others, 2016)
3. Genetic association with healthy behaviors (Liu, Rea-Sandin, Foerster, Fritsche, Brieger, Clark, Li, Pandit, Zajac, Abecasis & others, 2017).
4. Numerous correlational studies with non-cognitive outcomes using ICAR as a cognitive control.

Proprietary software

1. Just as ability tests were sold for a profit, so were statistical systems profit making.
2. The history of computing in the social sciences shows that although the major statistical systems originally were developed at universities to solve specific statistical problems, they soon were spun off into profit making ventures (Revelle, Elleman & Hall, 2020):
   - BMDP was developed for biomedical research
   - SAS® for agriculture research
   - SPSS for statistics in the social sciences (SPSS, 2008)
   - S^+ for general statistics
   - Systat for general statistics and graphics
3. Advanced statistics for structural equation modeling was also proprietary
   - LISREL
   - MPLUS

The open source revolution

1. In 1992 Ross Ihaka and Robert Gentleman, at the University of Otago in New Zealand, adapted S to work on Macs and called their resulting product, R.
2. R incorporated the list oriented language Scheme and emphasized object-oriented programming.
3. Most importantly, they shared the design specifications with other interested developers around the world
4. They intentionally did not copyright the code.

Thus was R born and become the amazing resource it is today (R Core Team, 2020)

The power of R

1. The real power of R is that because it is open source, it is extensible.
2. Anyone can contribute packages to the overall system. These packages are then distributed through CRAN.
3. R is a statistical system as well as an efficient object oriented programming language. Much of R is written in R.
4. That, and the power of the General Public License (GPL) and open source software movement has led to an amazing effect.
5. From the original functions in R and the ones written by the R Core Team (R Core Team, 2020), more than 16,734 packages have been contributed to CRAN, the Comprehensive R Archive Network, and at least more than 34,000 packages are available on GitHub.
6. R runs on most computing platforms (Unix and Unix likes, PCs, Macs, the web).

A sample of useful packages to analyze the big data of education

1. *psych* (Revelle, 2020a) is a "Swiss Army Knife" that can do many things for basic descriptive and inferential data analysis and psychometrics (e.g., EFA, IRT, α, β, ω_h, ω_t) but is not the best for any one thing. As of November, 2020, it has ≈ 160K monthly downloads with 7.5M total downloads.
2. *lavaan* (Rosseel, 2012) is the definitive package for Confirmatory Factor Analysis and Structural Equation Modeling. It has ≈37K downloads/month and 1.3M total downloads
3. *GPArotation* (Bernaards & Jennrich, 2005) is the go to package for factor/component rotations. It has ≈ 23K downloads/month and 788K total downloads.
4. *mirt* (Chalmers, 2012) is a powerful package for univariate and multivariate Item Response Theory analyses. It has ≈ 9K downloads/month and 315K total downloads.

Most packages come with 'vignettes' that describe how to use them

Vignettes for the *psych* package include:

1. An introduction to the psych package: part I
2. An introduction, part II: scale construction and psychometrics
3. How to use the psych package for factor analysis
4. Using R to score personality scales
5. Using R and the psych package to find ω
6. How to do mediation/moderation/regression analysis
7. How to install R and the *psych* package

Others have helpful web pages e.g., lavaan.

Open Source Code

1. All package and the base R code are open source and may be read, modified, and reused.
2. The 16K+ packages on CRAN and their associated functions can be adapted for your own use if you choose to do so.
3. R is open to all who want to contribute, is relatively easy to use and is growing. Providing R code in a manuscript or even creating a new package leads to the rapid adoption of new methods.
4. The advantages of open source is that when bugs are found (and they will be), a report to the package's author will lead to a documented fix to the code.
5. Packages keep getting added to CRAN, revisions are made to the base R roughly every six months, revisions are made to packages sporadically (e.g., psych about three times a year).
6. Always check the version number you are using.

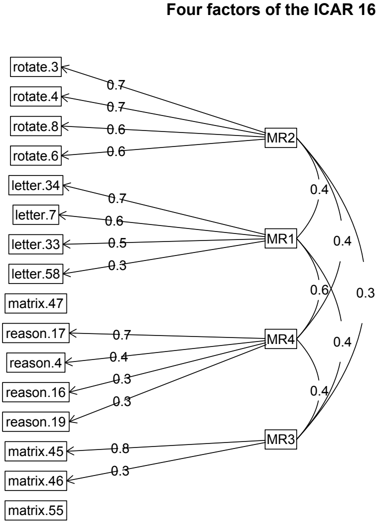
psych is intentionally easy to use; one to three lines do it

1. Find four factors for the ICAR sample test using the `fa` function,
2. Draw the solution using `fa.diagram`
3. Find a higher order solution using the `omega` function
4. Report various reliability statistics (Revelle & Condon, 2019)
5. Show the item information curves from an irt analysis of the ability data set.

R code

```
f4 <-fa(ability,4)#use the EFA function fa to find 4 factors
fa.diagram(f4,main="Four factors of the ICAR 16")#diplay them
om <- omega(ability,4) # Find a hierarchical factor solution
summary(om) #give the reliabilities
plot(om,sl=FALSE,main="Hierarchical solution of the ICAR 16")


icar.irt <- irt.fa(ability)  #show the item information
plot(icar.irt,type="test") #show the test information
```

Four factors for the ICAR sample test




```
#use the EFA function fa
          to find 4 factors
f4 <-fa(ability,4)

# Display results with
        fa.diagram

fa.diagram(f4,main="Four
        factors of the ICAR 16")
```

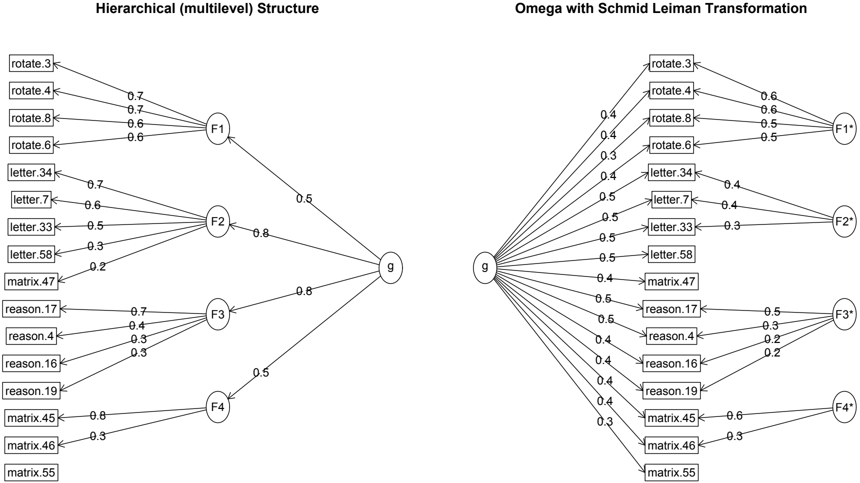
Reliability estimates for the ICAR 16 from the `omega` function

R code

```
om <- omega(ability,4)
summary(om)
```

```
omega(m = ability, nfactors = 4, main = "A Schmid Leiman solution")
Alpha:                  0.83
G.6:                    0.84
Omega Hierarchical:     0.66
Omega H asymptotic:     0.77
Omega Total             0.86

With eigenvalues of:
   g  F1*  F2*  F3*  F4*
3.05 1.31 0.47 0.40 0.53
The degrees of freedom for the model is 62  and the fit was  0.05
The number of observations was  1525  with Chi Square =  70.96  with prob <  0.2
The root mean square of the residuals is  0.01
The df corrected root mean square of the residuals is  0.02
RMSEA and the  0.9 confidence intervals are  0.01 0 0.019
BIC =  -383.48              Explained Common Variance of the general factor =  0.53

 Total, General and Subset omega for each subset
                                                g  F1*  F2*  F3*  F4*
Omega total for total scores and subscales    0.86 0.77 0.69 0.64 0.52
Omega general for total scores and subscales  0.66 0.24 0.52 0.47 0.27
Omega group for total scores and subscales    0.13 0.53 0.17 0.17 0.25
```

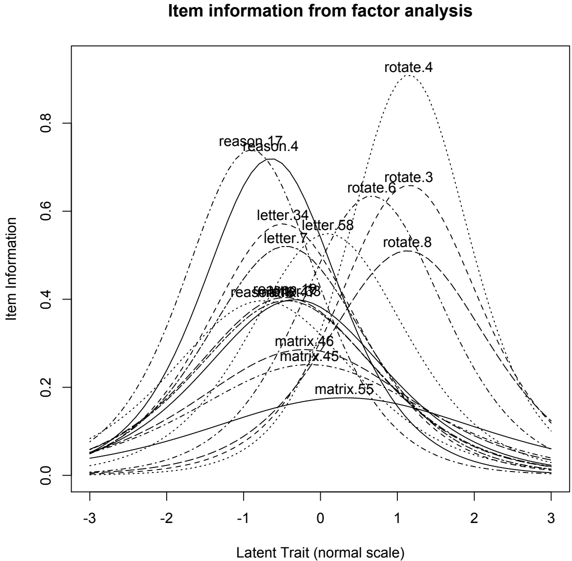
Hierarchical and SL transformations of the ICAR sample test




```
om <- omega(ability,4) #draws the Schmid Leiman solution
plot(om,sl=FALSE,main="Hierarchical solution of the ICAR 16")
```

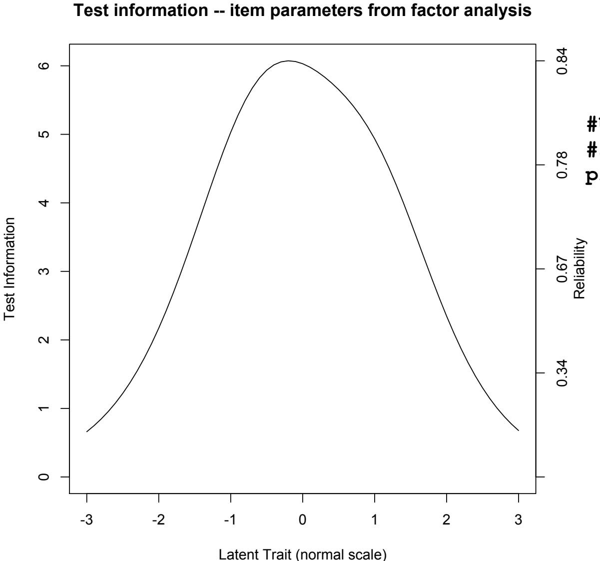
NORTHWESTERN UNIVERSITY

Item information from factor analysis for the ICAR sample test



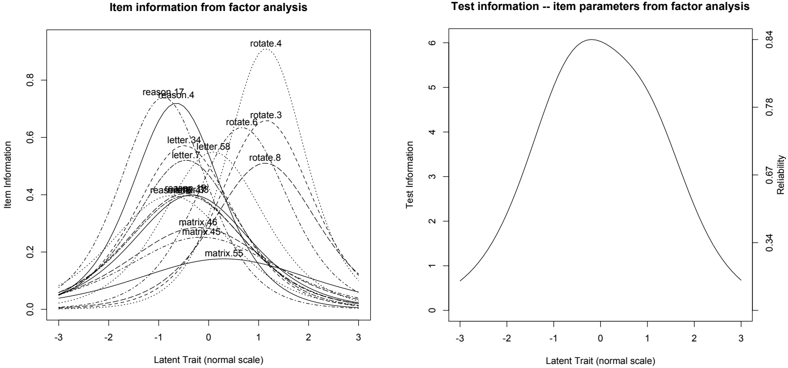

```
#Do a factor analysis of
 tetrachoric correlations
#Convert the output to
     IRT  format
#display the item
  information curves

icar.irt - irt.fa(ability)
```

Total test information (and reliabilty) varies by score




```
#Use the previous result
#show the test information
plot(icar.irt,type="test")
```

Item and test information curves for the ICAR sample test



IRT approach considers item difficulty and discrimination and be found from conventional factor analysis of the tetrachoric correlation matrix, with the addition of a difficulty parameter.

The Open Science Movement has a long history and a great future

1. Open science allows us to ask our questions in a manner that can be examined publicly.
2. Science is not just for the rarefied experts, it is for everyone.
3. Science will be trusted when it is not a secret shared among the elite but when all of mankind can contribute.
4. The ICAR and R are examples of open science.
5. The study of cognitive ability allows us to identify potential leaders as well as to to identify missing variables in our research.
6. With the International Cognitive Ability Resource the study of cognitive ability is now open to all investigators.
7. With open source software, e.g., R, computation and analysis is open to anyone who is willing to learn.

These slides and supplementary material are available at personality-project.org/sapa

Bernaards, C. & Jennrich, R. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*(5), 676–696.

Binet, A. & Simon, T. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'annee Psychologique*, *12*, 191–244 Binet, 1905 A. Binet, New methods for the diagnosis of the intellectual level of subnormals, L'Annee Psychologique 12 (1905), pp. 191–244 (Translated in 1916 by E. S. Kite in The Development of Intelligence in Children. Vineland, NJ: Publications of the Training School at Vineland.).

Bowman, M. L. (1989). Testing individual differences in ancient China. *American Psychologist*, *44*(3), 576–578.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29.

Condon, D. M. (2012). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. Master's thesis, Northwestern University, Evanston, Illinois, USA.

Condon, D. M. (2018). *The SAPA Personality Inventory: An empirically-derived, hierarchically-organized self-report personality assessment model*. PsyArXiv /sc4p9/.

Condon, D. M., Doebler, P., Holling, H., Gühne, D., Rust, J., Stillwell, D., Sun, L., Chan, F., Loe, A., & Revelle, W. (2014). International Cognitive Ability Resource. *https://icar-project.com*.

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. (2016). Selected ICAR data from the SAPA-Project: Development and initial validation of a public-domain measure. *Journal of Open Psychology Data*, *4*(1).

NORTHWESTERN UNIVERSITY

Damian, R. I., Spengler, M., Sutu, A., & Roberts, B. W. (2019). Sixteen going on sixty-six: A longitudinal study of personality stability and change across 50 years. *Journal of Personality and Social Psychology*, *117*(3), 674–695.

Deary, I. J. (2008). Why do intelligent people live longer? *Nature*, *456*(7219), 175–176.

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, *63*(1), 453–482. PMID: 21943169.

Dubois, P. H. (1947). The classification program report no. 2. Army air forces aviation psychology program research reports, Army Air Forces, Defense Documentation Center Defense Supply Agency.

Dworak, E. M., Revelle, W., Doebler, P., & Condon, D. M. (2020). Using the International Cognitive Ability Resource as an open source tool to explore individual differences in cognitive ability. *Personality and Individual Differences*.

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. London: Macmillan and Co.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79–132.

Gühne, D., Doebler, P., Condon, D. M., Luo, F., & Sun, L. (2020). Validity and reliability of automatically generated propositional reasoning items: A multilingual study of the challenges of verbal item generation. *European Journal of Psychological Assessment*.

Haier, R. J. (2016). *The neuroscience of intelligence*. Cambridge University Press.

Hunt, E. (2010). *Human intelligence*. Cambridge University Press.

Jensen, A. R. (2002). Galton's legacy to research on intelligence. *Journal of Biosocial Science*, *34*, 145–172.

Krieke, L. V. D., Jeronimus, B. F., Blaauw, F. J., Wanders, R. B., Emerencia, A. C., Schenk, H. M., Vos, S. D., Snippe, E., Wichers, M., Wigman, J. T., et al. (2016).

HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. *International journal of methods in psychiatric research*, *25*(2), 123–144.

Kuncel, N. R. & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, *315*(5815), 1080–1081.

Leon, M. R. & Revelle, W. (1985). Effects of anxiety on analogical reasoning: A test of three theoretical models. *Journal of Personality and Social Psychology*, *49*(5), 1302–1315.

Liebert, M. (2006). A public-domain assessment of music preferences as a function of personality and general intelligence. Honors Thesis. Department of Psychology, Northwestern University.

Liu, M., Rea-Sandin, G., Foerster, J., Fritsche, L., Brieger, K., Clark, C., Li, K., Pandit, A., Zajac, G., Abecasis, G. R., et al. (2017). Validating online measures of cognitive ability in genes for good, a genetic study of health and behavior. *Assessment*.

Loe, B., Sun, L., Simonfy, F., & Doebler, P. (2018). Evaluating an automated number series item generator using linear logistic test models. *Journal of Intelligence*, *6*(2), 20.

Loe, B. S. (2020). The development and validation of an online spatial network measure. *Assessment*, *27*(8), 1914–1927. PMID: 30674202.

Loe, B. S. & Rust, J. (2017). The perceptual maze test revisited: Evaluating the difficulty of automatically generated mazes. *Assessment*, *0*(0).

Lubinski, D. & Benbow, C. P. (2006). Study of Mathematically Precocious Youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, *1*(4), 316–345.

Lubinski, D., Benbow, C. P., Webb, R. M., & Bleske-Rechek, A. (2006). Tracking exceptional human capital over two decades. *Psychological Science*, *17*(3), 194–199.

Merz, Z. C., Lace, J. W., & Eisenstein, A. M. (2020). Examining broad intellectual abilities obtained within an mturk internet sample. *Current Psychology*, 1–9.

Murray, H. A., MacKinnon, D. W., Miller, J. G., Fiske, D. W., & Hanfmann, E. (1948). *Assessment of Men*. New York: Rinehart.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Revelle, W. (2020a). *psych: Procedures for Personality and Psychological Research*. https://CRAN.r-project.org/package=psych: Northwestern University, Evanston. R package version 2.0.9.

Revelle, W. (2020b). *psychTools: Tools to Accompany the Psych Package for Psychological Research*. https://CRAN.r-project.org/package=psychTools: Northwestern University, Evanston. R package version 2.0.8.

Revelle, W. & Condon, D. M. (2019). Reliability: from alpha to omega. *Psychological Assessment*, *31*(12), 1395–1411.

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., Dworak, E. M., & Condon, D. M. (2020). Cognitive ability in everyday life: the utility of open source measures. *Current Directions in Psychological Science*, *29*(4), 358–363.

Revelle, W., Elleman, L. G., & Hall, A. (2020). Statistical analyses and computer programming in personality. In P. J. Corr (Ed.), *Cambridge University Press Handbook of Personality* (pp. 495–534). Cambridge University 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.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Sackett, P. R. & Kuncel, N. R. (2018). Eight myths about standardized admissions testing. In B. W. Jack Buckley, Lynn Letukas (Ed.), *Measuring Success: Testing, Grades, and the Future of College Admissions* chapter 1, (pp. 13–38). Baltimore, MD: Johns Hopkins University Press.

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology*, *102*(3), 254 – 273.

Shipley, W. C. (2009). *Shipley-2: manual*. Western Psychological Services.

Sobkow, A., Olszewska, A., & Traczyk, J. (2020). Multiple numeric competencies predict decision outcomes beyond fluid intelligence and cognitive reflection. *Intelligence*, *80*, 101452.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology*, *15*(2), 201–292.

Spengler, M., Damian, R. I., & Roberts, B. W. (2018). How you behave in school predicts life success above and beyond family background, broad traits, and cognitive ability. *Journal of Personality and Social Psychology*, *114*(4), 600–636.

SPSS (2008). Version 17.0. *Chicago: SPSS Inc*.

Svenson, A., Guillen, P., et al. (2020). The links between intelligence, personality, and theory of mind in an adult sample. Economics Working Paper Series 10, University of Sydney.

Terman, L. M. (1916). *The measurement of intelligence*. Houghton Mifflin.

Terman, L. M. & Oden, M. (1947). *Genetic studies of genius*. Palo Alto, CA: Stanford University Press; Oxford University Press.

Terman, L. M. & Oden, M. (1959). *The gifted group at mid-life: Thirty-five years' follow-up of the superior child*, volume 5. Stanford Univ Pr.

Urbina, S. (2014). *Essentials of psychological testing*. John Wiley & Sons.

Yoakum, C. S. & Yerkes, R. M. (Eds.). (1920). *Army Mental Tests*. Henry Holt and Company.

Young, S. R. & Keith, T. Z. (2020). An examination of the convergent validity of the icar16 and wais-iv. *Journal of Psychoeducational Assessment*, *0*(0), 0734282920943455.

Zhang, H. (1988). Psychological measurement in China. *International Journal of Psychology*, *23*(1-6), 101–117.