

Exploring the persome: The power of the item in understanding personality structure

William Revelle*, Elizabeth M. Dworak*, David M. Condon

Abstract

We discuss methods of data collection and analysis that emphasize the power of individual personality items for predicting real world criteria (e.g., smoking, exercise, self rated health). These methods are borrowed by analogy from radio astronomy and human genomics. Synthetic Aperture Personality Assessment (SAPA) applies a matrix sampling procedure that synthesizes very large covariance matrices through the application of massively missing at random data collection. These large covariance matrices can be applied, in turn, in Persome Wide Association Studies (PWAS) to form personality prediction scores for particular criteria. We use two open source data sets (N=4,000 and 126,884 with 135 and 696 items respectively) for demonstrations of both of these procedures. We compare these procedures to the more traditional use of “Big 5” or a larger set of narrower factors (the “little 27”). We argue that there is more information at the item level than is used when aggregating items to form factorially derived scales.

Keywords: Persome, Persome Wide Association Studies, Synthetic Aperture Personality Assessment (SAPA), Massively Missing Completely at Random (MMCAR), Scale construction, Factor analysis, Item analysis; Open Source

The founding of the International Society for the Study of Individual Differences and its journal, *Personality and Individual Differences* were inspired by the eclectic interests of Hans Eysenck. Eysenck acted on the belief that the field of individual differences could benefit from many different approaches with sometimes contradictory results. As the editor of *PAID* he was happy to publish theories and results that were controversial, some in strong criticism of his own

*Correspondence concerning this article should be addressed to William Revelle, Department of Psychology, Northwestern University, Evanston IL, 60208. Telephone number: 847-491-7700. Email: revelle@northwestern.edu. This is the authors' version as submitted to PAID.

work, some that others would strongly criticize. His basic principal seemed to be that truth will out.

In this tradition, we introduce a procedure that is in direct contradiction to the tendency of most researchers to ignore the information that is available at the single personality item and to form higher level composites that are thought to reflect broad latent variables. We reverse this approach and emphasize the importance of the item information. We are not alone in this endeavor, for a few others have made similar suggestions ([McCrae, 2015](#), [Mottus, Kandler, Bleidorn, Riemann and McCrae, 2017](#), [Möttus, Sinick, A.Terracciano, Hřebíckova, Kandler and Jang, 2019](#)). Our contribution to this debate is to offer some new data collection and analytical procedures borrowed from other fields (specifically radio astronomy and genomics). We will also emphasize that our approach is based upon the open science framework that can be associated with the founding of the Royal Society which has as its motto “Nullius in verba” as a way of sharing ideas and findings in an open manner while relying on facts determined by experiment.

For those trained psychometrically, it is well known that forming item composites leads to increases in reliability ([Revelle and Condon, 2019](#)). The use of factor analysis is supposed to identify the latent variable accounting for the shared and reliable variance of the items. The remaining variance in the item is assume to reflect noise and in fact, once the trait is removed, the concept of local independence suggests that there is no meaningful variance left. This approach of discovering latent variables has a long and fruitful tradition going back to [Spearman \(1904\)](#) and ([Thurstone, 1933](#)) and used by Eysenk in his first explorations of the dimensions of personality ([Eysenck, 1944](#), [Eysenck and Himmelsweit, 1947](#)) as well as his contemporary Raymond Cattell ([Cattell, 1943, 1945, 1946](#)). However, we believe that by forming higher level constructs and ignoring the meaningful signals available at the item level, our field has been led astray.

Although early work on scale construction (e.g., [Strong, 1927](#), [Hathaway and McKinley, 1943](#), [Gough and Bradley, 1996](#)) emphasized the selection of items that predicted specific criteria, much of the past 50-80 years of personality scale construction has been concerned with identification of latent variables thought to measure the common variance of items. The use of factor analysis to identify these latent variables, and subsequent use of Structural Equation Models (SEM) which makes use of homogeneous scales in structural models has dominated the theoretical approaches to personality measurement.

In spite of the tendency to emphasize latent variables, there were some strong advocates of an external validity criteria for scale construction ([Jackson, 1970, 1971](#)) who continued to argue for an empirical and theory based approach to item

writing. Indeed, an influential paper suggested theoretical, external/empirical, and latent trait approaches did not differ in their average validities ([Hase and Goldberg, 1967](#)). However, a follow up to that article compared empirically constructed to latent variable methods in scale construction ([Goldberg, 1972](#)) and found that for hard to predict criteria, empirical methods were probably slightly superior and for easy to predict criteria, factorially homogeneous techniques were superior. Although factor analysis is clearly an empirical procedure, Goldberg made the distinction between empirical methods based upon the external validity of items and those based upon the internal structure of scales using factor analytic approaches. He labeled these two techniques as empirical versus factorial. This article can be seen as a followup to that work.

It has been known since 1910 ([Brown, 1910](#), [Spearman, 1910](#)) that combining items into scales leads to an increase in reliability. Given that most data are befuddled with error ([McNemar, 1946](#)) and psychological data are even more befuddled than most, forming item composites will partially compensate for this befuddlement. The concept of correcting for reliability of the measures when examining associations between measures ([Spearman, 1904](#)) requires stable estimates of reliability. With the sample sizes typically found in many studies for the past century, such stability could only be achieved at high levels of reliability. But, another way to achieve stability of estimates is to increase the sample size. When this is done, the benefits of item level analysis become more apparent.

But how to increase sample size and at the same time have a large item pool to take advantage of the power of the item? The solution was originally discussed by ([Lord, 1955](#)) in terms of sampling items as well as people, an analogous procedure to do so was developed in radio astronomy.

0.1. Synthetic Aperture Personality Assessment

The development of the telescope by Galileo Galilei revolutionized the way humans see the world. Galileo's original telescope was just eight power and after successive improvements, with his 30 power telescope he was able to detect what we now call the Gallilean moons of Jupiter but which he politically called the Medician moons after his sponsor (giving credit to sponsoring organizations such as the NSF, the MRC, or the NIH is not a new idea). Telescopes then and now are limited by the amount light that they capture. This is proportional to the cross section of the telescope and the refractor telescope of Galileo's first telescope was just 37 mm in diameter. His later telescopes were slightly larger, but were limited in their resolution by being refractor telescopes. Subsequently, Isaac Newton developed the reflector telescope which had the advantage that it could have a greater aperture and did not suffer from chromatic aberration. Modern

optical telescopes all follow Newton's model. Originally expressed in inches of diameter of a single mirror (e.g. the Mount Wilson 100 inch and the Palomar 200 inch) with the development of compound mirrors the sizes exploded to the largest of these (currently the Canary Islands observatory has the largest single mirror at 10.4 m diameter, telescopes with compound mirrors are being built in Chile with a 25m and 39M diameter and there is talk of a 30 meter telescope on Mona Kea in Hawaii).

Optical telescopes record visible light in a small part of the electromagnetic spectrum. Radio telescopes record a different part of the spectrum but more importantly for our purposes have taken a different approach to increasing the aperture. Rather than a single telescope of great diameter (think of the Arecibo radio telescope in Puerto Rico with a diameter of 305 meters) it is possible to synthetically integrate the signals from multiple telescopes spread out over several kilometers at one site or even spread around the world in which case the resolving power is effectively limited by the diameter of the earth.

Why have we taken this diversion to discuss radio astronomy? Because of the analogy of the synthetic aperture radio telescope to Synthetic Aperture Personality Assessment (SAPA). Just as the resolving power of a single telescope is limited by its diameter, so is the resolving power of a single personality questionnaire limited by the number of items in the questionnaire and the number of people taking the items. But if we could give many different forms of the questionnaire, with different items to different people, and then combine these signals, we have the resolving power of a much larger questionnaire. This is the basic idea behind SAPA.

SAPA is not a particularly new idea for the combining of scores from different sets of items has been advocated for quite a while ([Lord, 1955](#)). In order to increase the number of items given in national and international surveys such as the National Assessment of Educational Progress (NAEP) and the Program for International Student Assessment (PISA) balanced incomplete block designs are common where students are systematically given different forms ([Anderson, Lin, Treagust, Ross and Yore, 2007](#)). Extensions of balanced incomplete blocks are known as matrix sampling with planned missingness designs ([Graham, Taylor, Olchowski and Cumsille, 2006](#), [Little, Gorrall, Panko and Curtis, 2017](#), [Rhemtulla, Savalei and Little, 2016](#)) which share the concept of a design with blocks of items missing at random (MAR) and if the blocks are completely at random this is known as MCAR. The number of such blocks is typically not very large in order to allow for full information maximum likelihood (FIML) estimates of the data, However, this procedure to increase the number of items given in a

questionnaire is not particularly common in personality research and the usual tendency is to give the same short questionnaire to all participants (Bleidorn, Schönbrodt, Gebauer, Rentfrow, Potter and Gosling, 2016, Rentfrow, Gosling, Jokela, Kosinski and Potter, 2013, Soto and John, 2017). We have reported before (Revelle, Wilt and Rosenthal, 2010) using MCAR techniques to estimate temperament and ability scores from hundreds of items and then expanded the method to what we call Massively Missing Completely at Random (MMCAR) to estimate covariance structures of thousands of items with a level of missingness of up to 99% (Revelle, Condon, Wilt, French, Brown and Elleman, 2016).

The procedure is very simple and takes advantage of the web as well as people's interest in knowing themselves. At the web site (<https://sapa-project.org>) we give 50-250 items to each participant, with the items randomly sampled from about 6,000 items. (We started the original SAPA study by giving 50 items sampled from 120-140 item, but have gradually increased the number from which we sample to the current level of about 6,000.) Every subject gets feedback on their personality using a "Big 5" and a "little 27" (Condon, 2017) framework. This feedback seems to be compelling enough for users to tell others about the web site and, with the power of various search engines, attracts 50-100,000 people per year. Some other sites attract far more visitors, but give the same set of items to all participants.

But why care about the number of items if we already know there is a Giant 3 (Eysenck, 1994) or Big 5 (Digman, 1990, Goldberg, 1990)? Isn't enough to just give a short Big 5 inventory (e.g., Gosling, Rentfrow and Swann, 2003, Konstabel, Lönnqvist, Leikas, Velázquez, H, Verkasalo, and et al., 2017, Rammstedt and John, 2007) to see how personality correlates with criteria? No. It is better to consider large sets of items. To show the power of items we will report four studies using 5-27 scales and 135-696 items from two open source data sets ($N = 4,000$ and $126,884$) that we have made public. In the spirit of open science, we include the R (R Core Team, 2019) code that we use for the analyses, all of which are done using the open source *psych* package (Revelle, 2020a) for R.

We report our results in four different studies.

The first is a direct comparison of using conventional regression techniques to predict 10 different criteria for 4,000 participants from the *spi* dataset. These data are included in the *psychTools* package which accompanies the *psych* package (Revelle, 2020a) for the open source language and environment for statistical computing, R (R Core Team, 2019). The *spi* data set includes 135 items chosen from the 696 items shown by Condon (2014) to be common to more than 200 public domain scales. The second study applies a profile analysis to these same

items and criteria. The third and fourth studies generalize the first two studies to far more items (696) and far more subjects (126,884). Given our beliefs in the importance of replicable work and in the spirit of open science, the analyses and data in all four of these studies uses open source computer code (included in the appendix) as well as data that have been previously released for open analysis. Although we include 10 criteria in the first two studies and 19 in the second two, for all four studies, we emphasize those that have direct implications for health and are particularly relevant to the readers of this journal, e.g., the personality correlates of smoking.

Because of the large to very large sample sizes involved, we do not report the conventional “statistical significance” nor even the confidence intervals of our effects. We do report the largest standard error for the correlations for each study. We prefer to use cross validation of our effects to show their stability. Suffice it is to say that any effect we report differs from a null effect. Following the advice of [Funder and Ozer \(2019\)](#) and others we report effect sizes in terms of correlations rather than squared correlations.

1. Study 1: Regressions using spi data set

The first study uses data from the spi data set which is included in the *psychTools* package and includes 135 items (the SAPA Personality Inventory aka the SPI-135, [Condon, 2017](#)) as well as ten criteria. The items were first scored to form five (“Big 5”) scales, then 27 (“little 27”) scales. These 5, 27 and then all 135 items were then entered into regression models and cross validated. The 135 spi items were chosen from the 696 items shown by [Condon \(2014\)](#) to be common to more than 200 public domain scales. The 27 scales were derived by factoring the entire 696 item pool and are variously described as “facets”, “factors” and “scales”. The diminutive of “little 27” is to distinguish them from claims about “The Big 5”, the “Giant 3” or even the “General 1”.

1.1. Data: Variables and Subjects

The data for the first study come from 10 different criteria for 4,000 participants from the spi dataset which is included in the *psychTools* package ([Revelle, 2020b](#)). The *psychTools* package accompanies the *psych* package ([Revelle, 2020a](#)) for the open source language and environment for statistical computing, R ([R Core Team, 2019](#)). The data were collected as part of the larger SAPA project.

Using SAPA data contributed by about 126,000 visitors to our website (<https://SAPA-project.org>) David [Condon \(2014\)](#) developed a heterarchical frame-

work for assessing personality at three levels: The highest level has the familiar five factors that have been studied extensively in personality research since the 1980s – Conscientiousness, Agreeableness, Neuroticism, Openness, and Extraversion. The middle level has 27 factors that are considerably more narrow. These were derived based on administrations of 696 public-domain IPIP (Goldberg, 1999) items to about 126,000 participants. The lowest level consists of the 135 items thought to best reflect the five and 27 factors. Condon describes these scales as being “empirically-derived” because relatively little theory was used to select the number of factors in the hierarchy and the items in the scale for each factor (to be clear, he means relatively little personality theory though he relied on quite a lot of sampling and statistical theory). The procedures for developing these scales are discussed in the manual for the Sapa Personality Inventory (Condon, 2017) which includes the R code and instructions for downloading the original data from Dataverse.

The 10 criteria are: Age (in years, from 11 - 90). Sex (self reported biological sex, coded by the number of X chromosomes as 1 or 2), Health (self rated health on a 1-5 scale from poor to excellent), P1Edu and P2Edu are reported level of education for the participants parents, Education is respondent’s education (less than 12 years, High School graduate, currently in university, some university, associates degree, college degree, in graduate or professional school, with a graduate or professional degree. Wellness (rated as 1-2). Exercise (frequency of exercise from very rarely to more than 5 times/week. Smoking (never, not last year, less than once a month, less than once a week, 1-3 days per week, most days, up to 5 times a day, up to 20 times a day, and more than 20 times a day (coded as 1 to 9). Emergency Room visits (none, 1, 2, 3, more than 3 times) (coded as 1 to 4) The basic descriptive statistics are shown in Table 1.

Table 1: Descriptive statistics for the 10 criteria variable taken from the spi data set.

Variable	vars	n	mean	sd	median	trmmd	mad	min	max	range	skew	krtss	se
age	1	4000	26.90	11.49	23	25.02	7.41	11	90	79	1.45	1.80	0.18
sex	2	3946	1.60	0.49	2	1.62	0.00	1	2	1	-0.39	-1.85	0.01
health	3	3536	3.51	0.98	4	3.54	1.48	1	5	4	-0.25	-0.42	0.02
p1edu	4	3051	4.72	2.39	5	4.77	4.45	1	8	7	-0.11	-1.33	0.04
p2edu	5	2896	4.33	2.32	5	4.28	4.45	1	8	7	0.09	-1.33	0.04
education	6	3330	4.10	2.21	3	4.00	1.48	1	8	7	0.41	-1.04	0.04
wellness	7	3311	1.54	0.50	2	1.55	0.00	1	2	1	-0.17	-1.97	0.01
exer	8	3310	3.57	1.60	4	3.60	1.48	1	6	5	-0.35	-1.06	0.03
smoke	9	3348	2.19	2.04	1	1.70	0.00	1	9	8	1.83	2.19	0.04
ER	10	3347	1.16	0.48	1	1.03	0.00	1	4	3	3.42	12.74	0.01

1.2. Method

The 135 items in the spi data may be scored for 5 broad personality traits (the “Big 5”) using 70 of the items as well as 27 narrower factors with five items per scale (the “little 27”). The ω_h and ω_{total} reliabilities, as well as α and the scale intercorrelations for the Big 5 are shown in Table 2. (See [Revelle and Condon, 2019](#), for a discussion of these and other coefficients.)

Table 2: Reliabilities and correlations of the Big 5 scales from the spi data set. α reliability is on the diagonal of the correlations.

Variable	ω_{total}	ω_h	Agree	Consc	Neuro	Extra	Open
Agreeableness	0.91	0.61	0.87				
Conscientiousness	0.89	0.61	0.24	0.86			
Neuroticism	0.93	0.71	-0.12	-0.19	0.90		
Extraversion	0.92	0.70	0.23	0.07	-0.20	0.89	
Openness	0.88	0.72	0.00	0.01	-0.12	0.13	0.84

Because ω statistics are not meaningful for the short 5 item scales of the little 27, we just summarize the α values which ranged from .67 (Easy goingness) to .90 (Well being) with a median and mean of .82.

Using the Big 5, and little 27 as well as all the 135 items, multiple regressions were conducted to predict each criteria from the Big5 scores. To eliminate the effect of capitalizing on chance, which is always a problem in multiple regression, and particularly problematic as the number of predictors increases, we randomly split the sample into two equal parts, and then developed the regression models on the first half and then cross validated on the second half. Such cross validation is important, for although most multiple regression procedures also predict the shrunken values of the regression, these seem to over estimates of the observed cross validated values. We compare cross validated regressions fousing the Big 5, the little 27, and all 135 items (Figure 1).

In addition to standard regression, we also applied a newly developed algorithm, `bestScales` ([Elleman, McDougald, Revelle and Condon, 2020](#)) which is included in the *psych* package. `bestScales` is an empirical scale construction procedure which identifies those items most correlated with a specific criterion. This process is repeated k times (so called k-fold cross validation) and the final scale is chosen from those items that identified in all the folds. Although not quite as powerful as standard machine leaning algorithms such as Lasso regression ([Tibshirani, 2011](#)) or the elastic net ([Zou and Hastie, 2005](#)), it is specifically designed to work with high levels of missing data (e.g. SAPA like data). One of the advantages of `bestScales` is that it allows identification of the best unit

weighed items predicting any particular criterion. We include these results as an additional line in Figure 1.

It is tempting to try to interpret the items that have significant regression weights in these models. When doing so, we must remember that regression weights reflect the independent contributions of each variable, with the effect of the other variables partialled out. We prefer to show the items chosen by the `bestScales` algorithm as these are chosen for their zero order correlations, rather than their regression weights (Table 3).

2. Study 2: Profile analysis yields additional information

2.1. GWAS like technique

To continue our science by analogy approach, we consider the similarity of single items in personality research to the Single Nucleotide Polymorphisms (SNPs) of genomic research. One of the exciting advances in modern genetics is the move from emphasizing single genes to studying the entire genome. Genome Wide Association Studies (GWAS) use large samples to detect the genetic effect of very small signals of individual SNPs on complex phenotypes. Because the individual signal is so small (with correlations of a SNP with a phenotype of $< .01$), very large samples are necessary to have enough power to detect the reliable effects above the background noise. Because 1,000s to 10,000s to 100,000s SNPs are being examined simultaneously corrections for multiple comparisons need to be made and the associated probability values typically used are 10^{-4} or even 10^{-5} and expressed in log units.

A graphic showing GWAS effects plots the effect size (and the $-\log$ of the probability) of SNPs within and between the chromosomes. Because their shape resembles that of a skyline, these are known as “Manhattan Plots”. We can do the same for the correlations of the individual items with various behavioral phenotypes grouped by personality scale (Figure 2). For the sake of simplicity, we show the correlations of the 135 SPI items with three criteria (self rated health, exercise, and smoking) for the 4,000 participants in the `spi` dataset. The top row of the plot shows the raw correlations, the bottom row $-\log p$ of the correlations. The probability values are corrected for multiple comparisons using the Holm correction (Holm, 1979) which is a more power test than the more typical Bonferroni.

What is obvious from this figure is that different items and different scales have different patterns of association with the different criteria. It is also apparent that the best items from some of the little 27 are better than the items that form the Big5 scales.

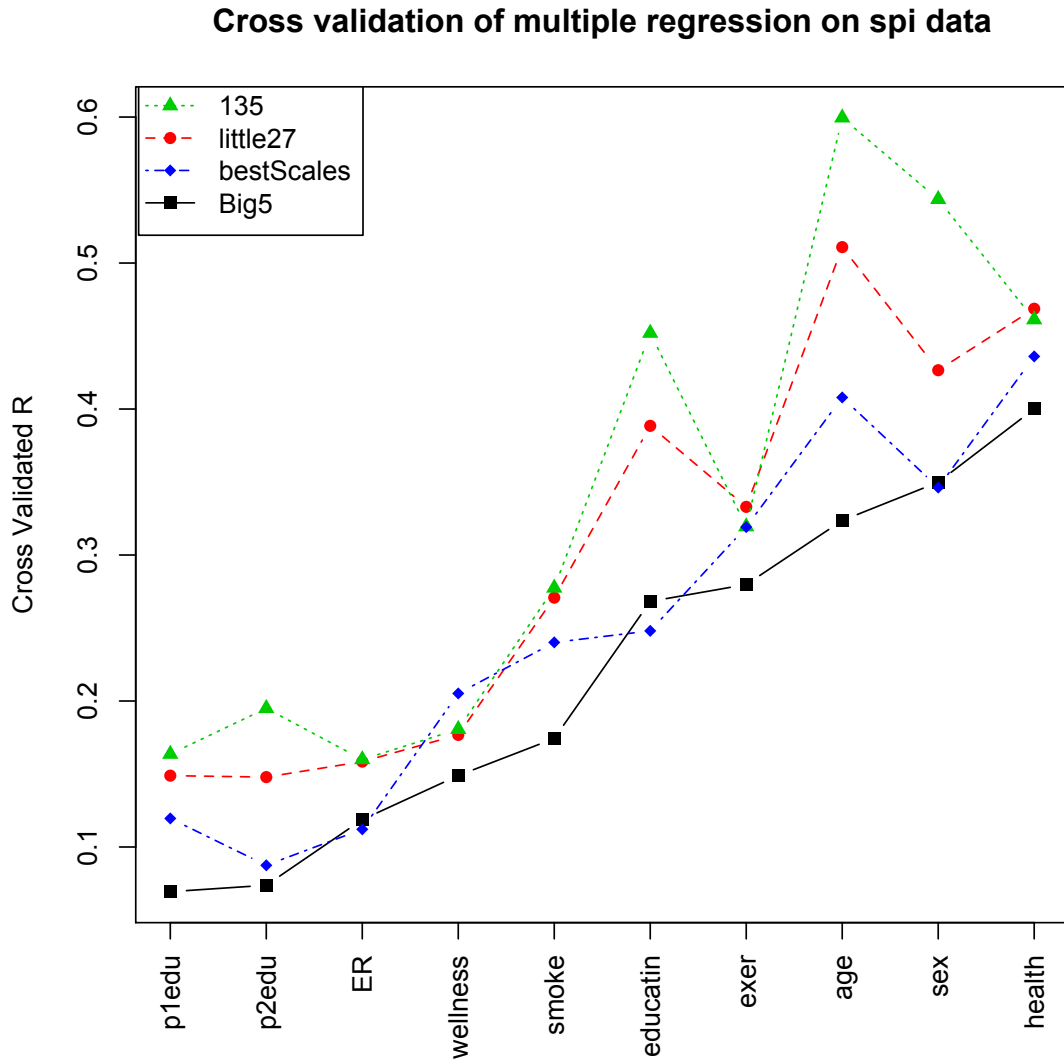


Figure 1: Cross validated correlations predictions 10 different criteria. Four methods of predictions are used. Big 5 were regression based models using the 5 Big 5 scales. Similarly, the little 27 and the 135 items represent cross validated multiple regression models. The “best scales” were found by adding up unit weighted scores based upon those items identified using the bestScales function.

Table 3: Zero order correlations of items from the spi that predict smoking and health. The composites of these items had cross validated predictive validities of .24 (smoking) and .44 (health). Items were identified using `bestScales`. Item numbers are from the list of all SAPA items. Mean correlations are the mean values across all 10 folds of the k-fold derivation samples. Standard deviations are the standard deviation across the 10 fold replications.

Items predicting smoking form a scale with a cross validated prediction of .24			
Variable	mean r	sd r	Item Content
q_1461	-0.24	0.01	Never spend more than I can afford.
q_1867	-0.20	0.01	Try to follow the rules.
q_1609	0.19	0.01	Rebel against authority.
q_1173	0.17	0.01	Jump into things without thinking.
q_1624	-0.17	0.01	Respect authority.
q_369	-0.16	0.01	Believe that laws should be strictly enforced.
q_56	-0.16	0.01	Am able to control my cravings.
q_35	0.16	0.01	Act without thinking.
q_1462	-0.15	0.01	Never splurge.
q_1424	0.15	0.01	Make rash decisions.
q_736	-0.15	0.01	Easily resist temptations.
q_598	0.14	0.01	Do crazy things.
q_1590	-0.13	0.01	Rarely overindulge.
q_1452	0.13	0.01	Neglect my duties.
q_2765	-0.12	0.01	Am happy with my life.

Items predicting self rated health form a scale with a cross validated prediction of .44			
Variable	mean r	sd r	Item Content
q_820	0.36	0.01	Feel comfortable with myself.
q_811	-0.35	0.00	Feel a sense of worthlessness or hopelessness.
q_2765	0.35	0.00	Am happy with my life.
q_578	-0.34	0.01	Dislike myself.
q_1371	0.31	0.01	Love life.
q_56	0.28	0.01	Am able to control my cravings.
q_1505	-0.28	0.01	Panic easily.
q_808	-0.27	0.01	Fear for the worst.
q_4249	-0.27	0.01	Would call myself a nervous person.
q_1452	-0.24	0.01	Neglect my duties.
q_979	-0.24	0.01	Get overwhelmed by emotions.
q_39	0.24	0.01	Adjust easily.
q_4252	-0.24	0.01	Am a worrier.
q_1444	-0.23	0.00	Need a push to get started.
q_1024	-0.23	0.01	Hang around doing nothing.
q_1840	0.23	0.01	Think that my moods dont change more than most peoples do.
q_1989	-0.22	0.01	Worry about things.
q_1052	-0.21	0.01	Have a slow pace to my life.
q_254	0.21	0.01	Am skilled in handling social situations.

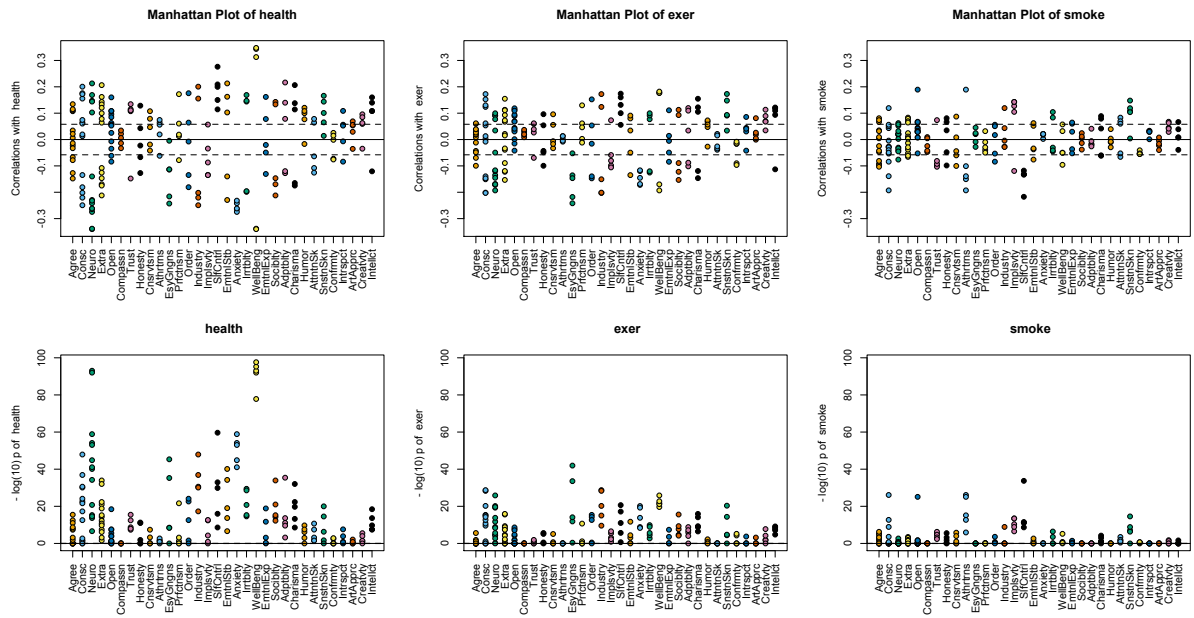


Figure 2: A ‘Manhattan’ plot of the item correlations with three criteria: health, exercise, and smoking organized by scale. The top panel shows the raw correlations for the items, the bottom panels show the log of the probability of the correlations. Items are organized by those forming the Big 5 scales (the first five columns in each panel, and then items in the little 27).

2.2. *Benefits of the item level approach*

One of the interesting techniques of GWAS is the use of genetic correlations (Nagel, Watanabe, Stringer, Posthuma and Van Der Sluis, 2018). These are the correlations found by examining the similarity of profiles across the genome of two phenotypic items. Although the phenotypic correlations between neuroticism items ranged from .17 to .54, when the correlation are taken between the profiles they ranged from .28 to .91 (Nagel et al., 2018). These correlations may be used to identify clusters of items showing similar genetic effects.

We apply this technique to find persome correlations, which are just the correlations of the profiles of correlations across items. Doing this leads to two sets of correlations: phenotypic correlations (the normal correlation of two criteria) and persome correlations (the correlation of the profiles across items). We show this in Figure 3: the lower off diagonal shows the phenotypic correlations, the upper off diagonal the persome correlations. The difference is clear: the persome correlations are much larger and show more clustering in their structure. For example, although the phenotypic correlation of smoking with emergency room visits is just .08, the correlation of what predicts smoking with what predicts emergency visits is .49. Similarly, while the correlation between exercise and self reported wellness is .15, the profile correlations of the patterns is .67.

It is important to note that although using 4,000 participants, the profiles are based upon the correlations across 135 items and thus the standard error for these correlations are $\sigma_r = \sqrt{\frac{1-r^2}{133}} < .087$. The standard errors of the phenotypic correlations are based upon the observed correlations and the sample size and are $< .016$

3. Study 3: Regressions with more items and more subjects

Using a second set of open source data, we analyze results for 126,884 subjects on 696 items as well as 19 criteria. The data were collected between April 2014 through February, 2017 as part of the SAPA project and may be downloaded from DataVerse (Condon and Revelle, 2015, Condon, Roney and Revelle, 2017a,b). As is our normal procedure, these data were collected using the MMCAR technique. Thus, of the 241,860 correlations ($696*695/2$) the median number of observations per pair was 2,704 with a range from 1,596 to 7,452 . Although using far more subjects than the spi data reported in studies 1 and 2, the SAPA data set has fewer observations per pairwise correlation (2,704) versus the complete data used in spi (4,000).

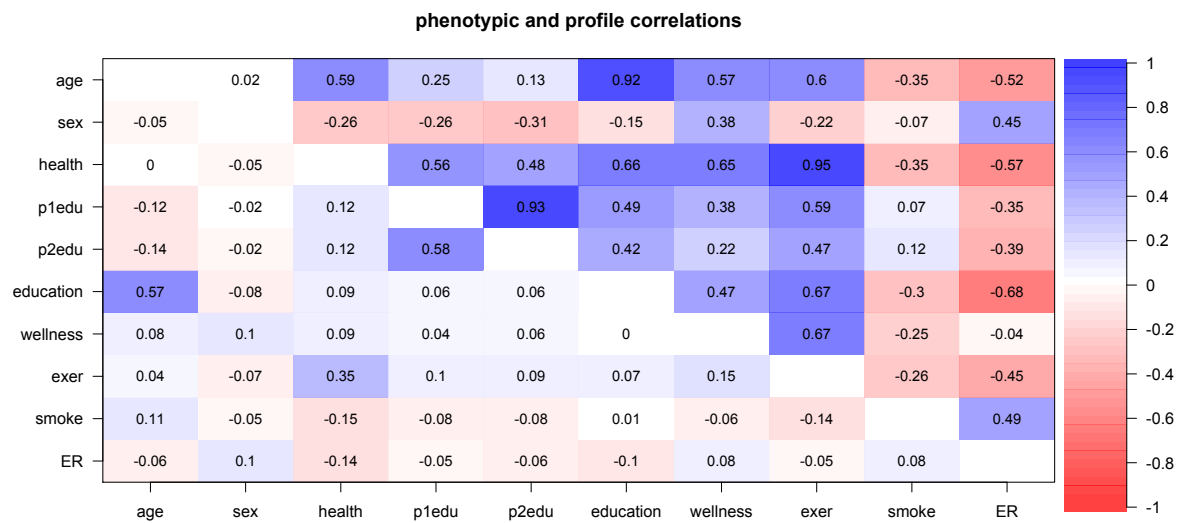


Figure 3: Phenotypic correlations between 10 criteria (lower off diagonal) and persome wide profile correlations (upper off diagonal). The sample is the 4,000 participants in the spi dataset. The standard errors of the phenotypic correlations are $< .016$ and the profile correlation are $< .087$.

3.1. Criteria used in the SAPA data set

In addition to the criteria used in the spi data set, are relationship and marital status, height, weight, Body Mass Index (BMI), job status, occupational prestige (for those with jobs), estimated income (based upon occupation), the occupational prestige and estimated income of both parents. Of these 19 criteria, seven overlap with those of the spi data set.

3.2. Regressions and bestScales for 5, 27, 135 and 696 items

Multiple regressions for each criteria were done using half the sample for the Big 5 scores, the little 27 scores, and regressions using the 135 spi items. In addition, the bestScales solutions for the 19 criteria was also found. All four of these solutions were then cross validated using the second half of the sample. We organize the results in terms of the Big 5 regressions (Figure 4).

Comparing the cross validated solutions for the complete data from the spi reported in Study 1 versus the much sparser data in the SAPA analyzed here are several interesting findings (Figures 1, 4). With the complete data of the spi data set, regressions using all 135 items had the largest cross validated values for 8 of the ten criteria, and were functionally tied with the little 27 regressions for the other 2 (Figure 1). The bestScales approach did not do as well as the little 27 regressions, but did do better than the Big 5 regressions. However, with a larger item pool and more missingness, the consistently best cross validated predictors were found by using a bestScales approach and the regressions for the 135 spi items did not do noticeably better (and sometimes noticeably worse) than the Big 5 regressions (Figure 4).

A strength of the empirical approach to scale construction (e.g., bestScales) is that it can identify the best items to predict particular criteria. A weakness is that it is completely mindless empiricism. For instance, the best scale to predict height has a cross validated value of .35. But the items that are most related to height are “panic easily”, “get overwhelmed by emotions” and “Am a worrier”. These non-sensical items are also the most correlated with gender. What the empirical scale construction technique is identifying is that women are shorter than men (with a correlation of -.65). But this is not just a fault of choosing the best items, for the regression weights for the Big 5 or the little 27 that best predict gender are almost exactly the opposite of those that predict height. The largest β weight for gender is Neuroticism (.18) as it is for height (-.14).

Cross validation of multiple regression on sapa data

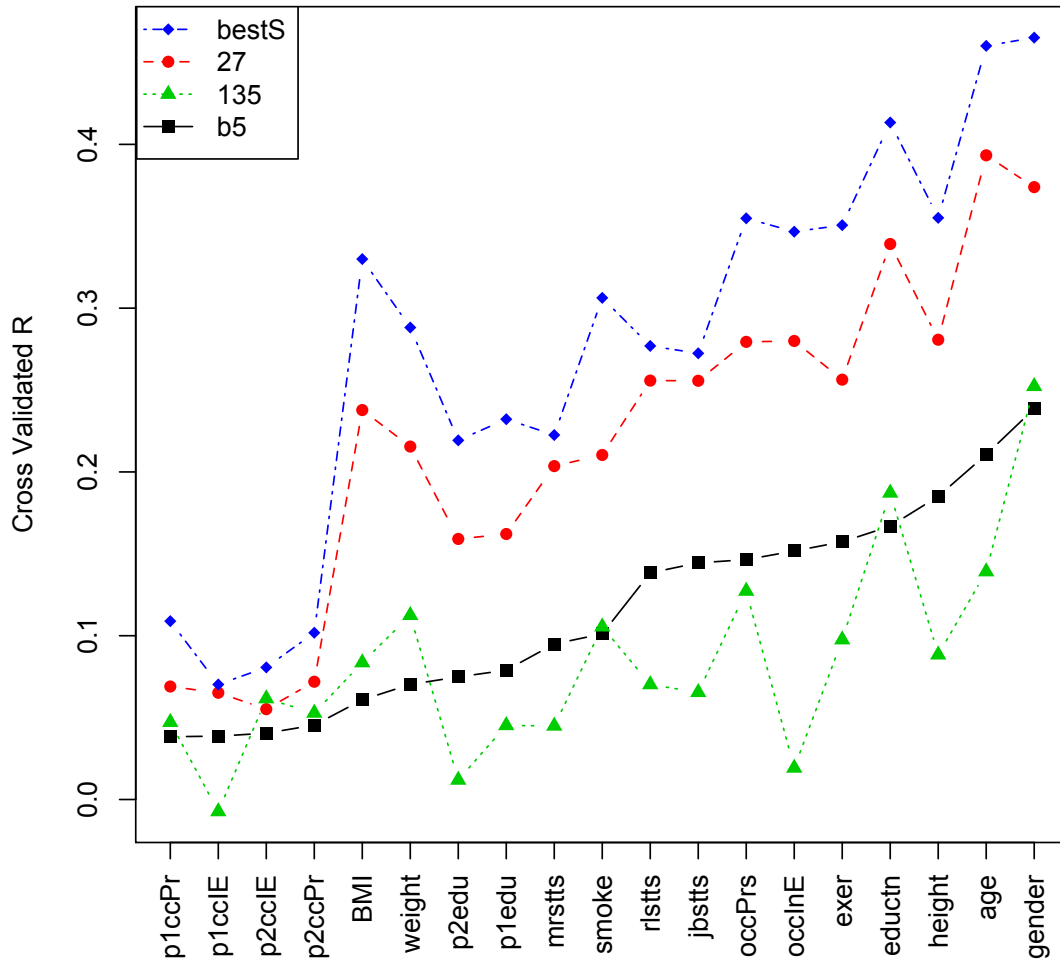


Figure 4: Predicting 19 critieria from the SAPA set. The lines show the cross validated regression or fits for regressions using the Big 5 scores, the little 27 scores, and the 135 items from the spi. Also shown are the cross validated values for the solution the bestScales function.

4. Study 4: Profile correlations using 696 items

Just as we could compare the phenotypic and persome profile correlations for the 10 criteria and 4,000 subjects of the spi data set, so we can compare the phenotypic and persome profile correlations for 19 criteria and 126,884 subjects. The profiles are based upon the correlations across 696 items and thus the standard error for these correlations are $\sigma_r = \sqrt{\frac{1-r^2}{694}} < .037$. The standard error of the phenotypic correlation based upon the sample size and the correlation and is $< .003$.

Comparing Figures 3 and 5 we see that the pattern of larger correlations and clearer cluster structure that we saw in Study 2 (Figure 3 is repeated with the larger sample and the increase in the number of criteria (Figure 5). The clearest cluster reflects parental education and occupation and a second clear cluster is the participants education and job status. (Education is highly correlated in this analysis with age, for younger participants have not had the opportunity to achieve higher levels of education. Although we control for age when predicting level of education for personality and ability measures in other studies, for the purposes of this demonstration, we have not done so.)

5. Study 5: Validating profiles across samples

One of the powerful applications of GWAS is the development of genetic propensity scores. These can be derived from large samples and then applied to much smaller samples. We do this here using the 135 items from our larger sample and then applying these profiles to the seven identical criteria in the spi dataset. Of the 19 criteria in the SAPA data set, seven of them appear in the spi data set. We identify the correlations in the SAPA data set for the overlapping criteria with the 135 items common to both data sets. These correlations may be used, in turn, to predict the criteria in the spi data. When we do this we find that the profile scores (basically weighting individual items by their zero-order correlations) are just slightly better than using the bestScales approach (which unit weights the best 20 items) and superior to regressions using the Big 5 factors. Profile scores are slightly worse than the regressions using the little 27 (Figure 6). Just as we saw for the full SAPA data, regressions using all 135 items were terrible, reflecting the instability of beta weights when using too many predictors.

The profile scores are the logical equivalent to genetic propensity scores and could be labeled as personalty propensity scores. For the important criterion of smoking behavior, the multiple R from the Big 5 was .18, while the optimal

Phenotypic (lower) and Profile (upper) correlations

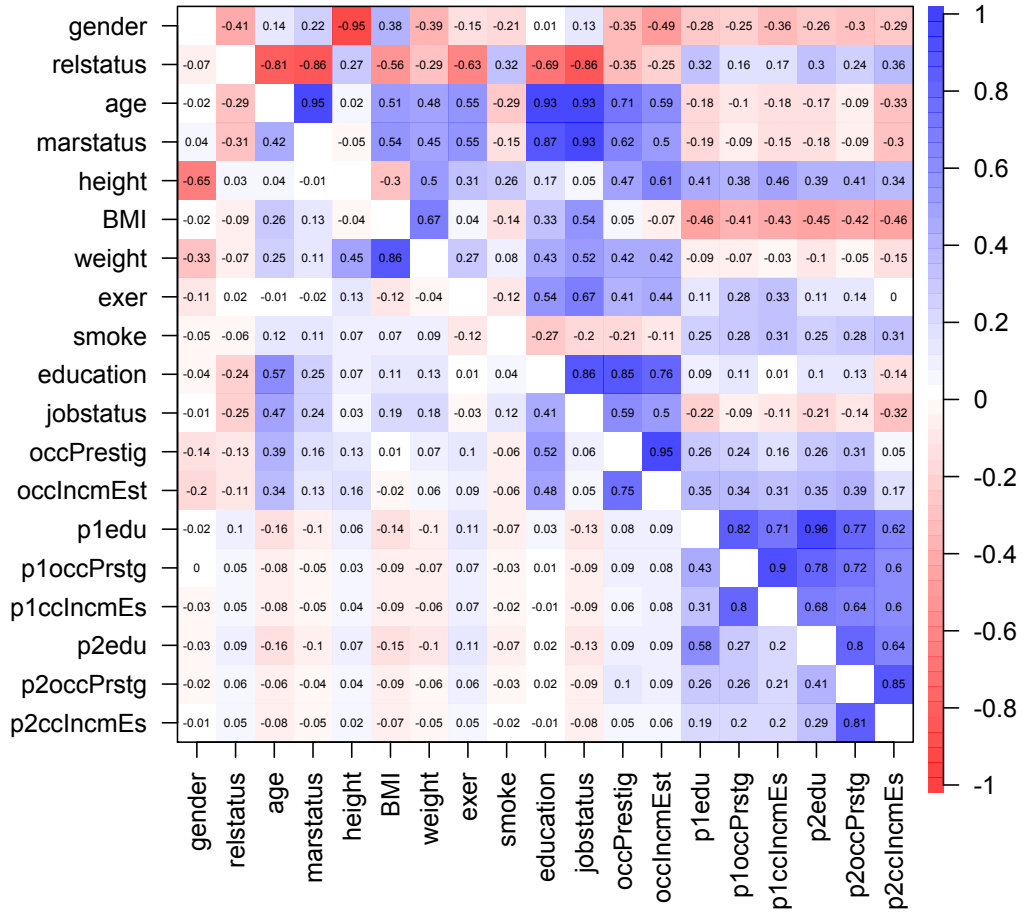


Figure 5: The phenotypic correlations (lower off diagonal) and persome profile correlations (upper off diagonal) of 19 criteria variables. The persome profiles are based upon 696 items. The profiles are based upon the correlations across 696 items and thus the standard error for these correlations are $\sigma_r = \sqrt{\frac{1-r^2}{694}} < .037$. The standard error of the phenotypic correlation based upon the sample size and the correlation and is $< .003$.

linear combination (regression weights) of 27 predictors had a multiple R of .29 and the personality propensity score and the bestScales approach were both .27 (Table 4).

Table 4: Predicting the spi criteria from the SAPA data set. Big 5 regressions are uniformly less than using the item information, either by profiles or best scales. Prediction models based upon the regression of all 135 items were with one exception inferior to all other models. Regression weights for the Little 27 had slightly higher validities.

Variable	Big 5	Little 27	135.R.	Best Scales.	Profiles.
p1edu	0.07	0.14	0.03	0.15	0.13
p2edu	0.09	0.17	0.06	0.15	0.16
smoke	0.18	0.29	0.24	0.27	0.27
exer	0.25	0.33	0.04	0.31	0.31
education	0.27	0.39	0.13	0.36	0.36
age	0.30	0.47	0.20	0.39	0.40
gender	0.35	0.42	0.06	0.38	0.42

6. Discussion and future directions

In this paper we have argued that there is more information at the item level than is used when aggregating items to form factorially derived scales. This is an old argument (Goldberg, 1972) that needs to be reconsidered. Factorially based scales are useful when limited in the number of items available, or with smaller sample sizes. But with the power associated with large sample sizes now available, it is time to revisit the use of items. Taking advantage of techniques analogous to those of radio astronomy and genomics allows us to improve our prediction of real world criteria.

In this paper we addressed the use of items that are more conventional in personality measurement. But following the tradition of the broader field of individual differences as seen in the pages of *Personality and Individual Differences* or at the meetings of the International Society for the Study of Individual Differences we are in the process of collecting a broader set of items that includes interests as well as cognitive abilities. We believe that by routinely measuring temperament, abilities, and interests and using the profile techniques discussed in this paper, we will have a better understanding of individual differences in personality.

Cross validation of multiple regression from SAPA on spi data

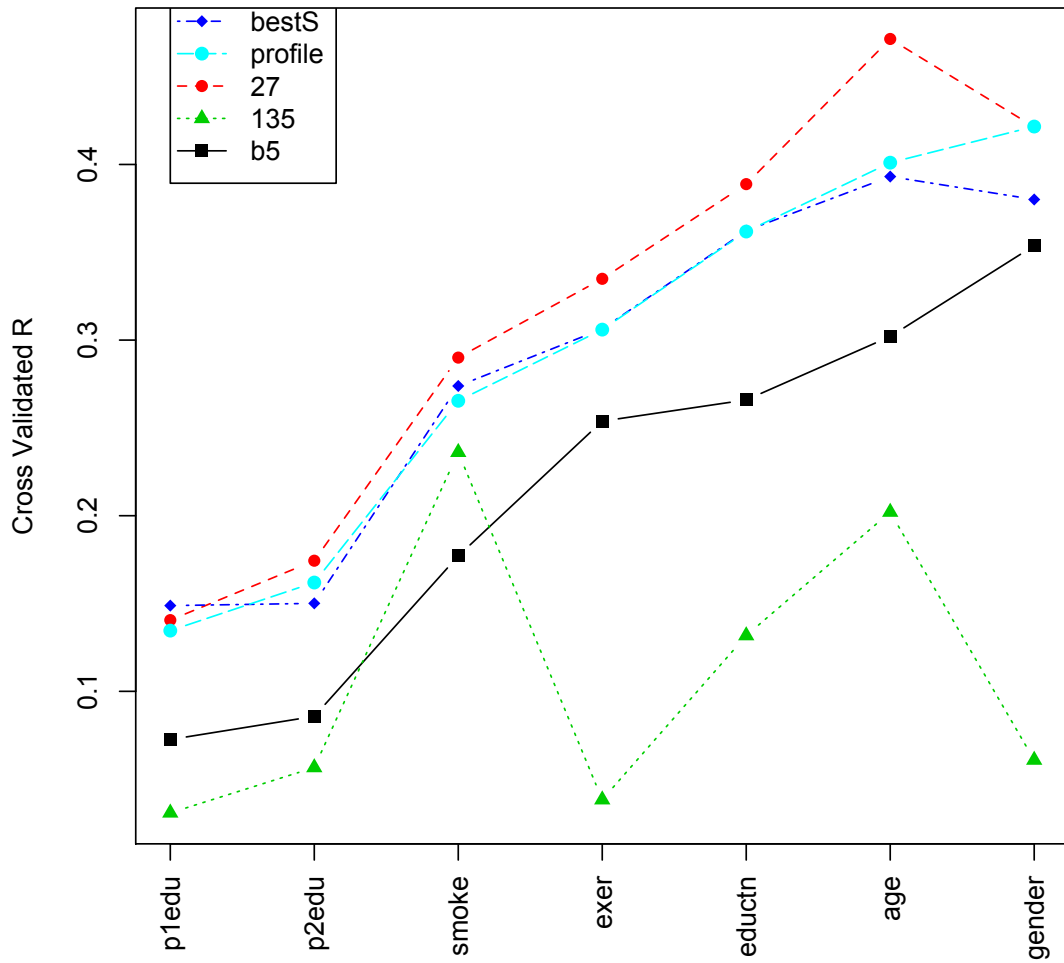


Figure 6: Predicting 7 criteria for the spi data from the larger SAPA set. The lines show the cross validated regression or fits for regressions using the Big 5 scores, the little 27 scores, and the 135 items from the spi. Also shown are the values for the solution the bestScales function as well as the profile scores.

References

- 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, 591–614. doi:[10.1007/s10763-007-9090-y](https://doi.org/10.1007/s10763-007-9090-y).
- Bleidorn, W., Schönbrodt, F., Gebauer, J.E., Rentfrow, P.J., Potter, J., Gosling, S.D., 2016. To live among like-minded others: Exploring the links between person-city personality fit and self-esteem. *Psychological Science* 27, 419–427. doi:[10.1177/0956797615627133](https://doi.org/10.1177/0956797615627133).
- Brown, W., 1910. Some experimental results in the correlation of mental abilities. *British Journal of Psychology* 3, 296–322. doi:[10.1111/j.2044-8295.1910.tb00207.x](https://doi.org/10.1111/j.2044-8295.1910.tb00207.x).
- Cattell, R.B., 1943. The description of personality. I. Foundations of trait measurement. *Psychological Review* 50, 559–594. doi:[10.1037/h0057276](https://doi.org/10.1037/h0057276).
- Cattell, R.B., 1945. The description of personality: Principles and findings in a factor analysis. *The American Journal of Psychology* 58, 69–90. doi:[10.2307/1417576](https://doi.org/10.2307/1417576).
- Cattell, R.B., 1946. Description and measurement of personality. World Book Company, Oxford, England.
- Condon, D.M., 2014. An organizational framework for the psychological individual differences: Integrating the affective, cognitive, and conative domains. Ph.D. thesis. Northwestern University.
- Condon, D.M., 2017. The SAPA Personality Inventory: An empirically-derived, hierarchically-organized self-report personality assessment model. Technical Report. Northwestern University. URL: <https://sapa-project.org/research/SPI/SPIdevelopment.pdf>.
- Condon, D.M., Revelle, W., 2015. Selected personality data from the SAPA-Project: 08dec2013 to 26jul2014. Harvard Dataverse doi:[10.7910/DVN/SD7SVE](https://doi.org/10.7910/DVN/SD7SVE).
- Condon, D.M., Roney, E., Revelle, W., 2017a. Selected personality data from the sapa-project: 22dec2015 to 07feb2017. [48,350 participant data file and codebook]. Harvard Dataverse doi:[10.7910/DVN/TZJGAT](https://doi.org/10.7910/DVN/TZJGAT).
- Condon, D.M., Roney, E., Revelle, W., 2017b. Selected personality data from the sapa-project: 26jul2014 to 22dec2015. [54,855 participant data file and codebook]. Harvard Dataverse doi:[10.7910/DVN/GU70EV](https://doi.org/10.7910/DVN/GU70EV).
- Digman, J.M., 1990. Personality structure: Emergence of the five-factor model. *Annual Review of Psychology* 41, 417–440. doi:[10.1146/annurev.ps.41.020190.002221](https://doi.org/10.1146/annurev.ps.41.020190.002221).
- Elleman, L.G., McDougald, S., Revelle, W., Condon, D., 2020. That takes the biscuit: A comparative study of predictive accuracy and parsimony of four statistical learning techniques in personality data, with data missingness conditions. *European Journal of Psychological Assessment* (in press).
- Eysenck, H.J., 1944. Types of personality: a factorial study of seven hundred neurotics. *The British Journal of Psychiatry* 90, 851–861.
- Eysenck, H.J., 1994. The big five or the giant three: Criteria for a paradigm, in: Halverson, C.F., Kohnstamm, G.A., Martin, R.P. (Eds.), *The developing structure of temperament and personality from infancy to adulthood*. Lawrence Erlbaum Associates, Hillsdale, N.J., pp. 37–51.
- Eysenck, H.J., Himmelweit, H.T., 1947. Dimensions of personality; a record of research carried out in collaboration with H.T. Himmelweit [and others]. Routledge & Kegan Paul, London.
- Funder, D.C., Ozer, D.J., 2019. Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science* 2, 156–168.

- Goldberg, L.R., 1972. Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monographs*. No 72-2 7.
- Goldberg, L.R., 1990. An alternative “description of personality”: The big-five factor structure. *Journal of Personality and Social Psychology* 59, 1216–1229. doi:[10.1037/0022-3514.59.6.1216](https://doi.org/10.1037/0022-3514.59.6.1216).
- Goldberg, L.R., 1999. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models, in: Mervielde, I., Deary, I., De Fruyt, F., Ostendorf, F. (Eds.), *Personality psychology in Europe*. Tilburg University Press, Tilburg, The Netherlands. volume 7, pp. 7–28.
- 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, 504 – 528. doi:[10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1).
- Gough, H.G., Bradley, P., 1996. *Manual for the California Psychological Inventory*.
- Graham, J.W., Taylor, B.J., Olchowski, A.E., Cumsille, P.E., 2006. Planned missing data designs in psychological research. *Psychological Methods* 11, 323 – 343.
- Hase, H.D., Goldberg, L.R., 1967. Comparative validity of different strategies of constructing personality inventory scales. *Psychological Bulletin* 67, 231–248.
- Hathaway, S., McKinley, J., 1943. *Manual for administering and scoring the MMPI*.
- Holm, S., 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, pp. 65–70.
- Jackson, D.N., 1970. A sequential system for personality scale development, Elsevier. volume 2 of *Current Topics in Clinical and Community Psychology*, pp. 61 – 96. doi:[10.1016/B978-0-12-153502-5.50008-4](https://doi.org/10.1016/B978-0-12-153502-5.50008-4).
- Jackson, D.N., 1971. The dynamics of structured personality tests: 1971. *Psychological Review* 78, 229 – 248. doi:[10.1037/h0030852](https://doi.org/10.1037/h0030852).
- Konstabel, K., Lönnqvist, J.E., Leikas, S., Velázquez, R.G., H, H.Q., Verkasalo, M., , et al., 2017. Measuring single constructs by single items: Constructing an even shorter version of the “short five” personality inventory. *PLoS ONE* 12, e0182714. doi:[10.1371/journal.pone.0182714](https://doi.org/10.1371/journal.pone.0182714).
- Little, T.D., Gorrall, B.K., Panko, P., Curtis, J.D., 2017. Modern practices to improve human development research. *Research in Human Development* 14, 338–349. doi:[10.1080/15427609.2017.1370967](https://doi.org/10.1080/15427609.2017.1370967).
- Lord, F.M., 1955. Sampling fluctuations resulting from the sampling of test items. *Psychometrika* 20, 1–22. doi:[10.1007/BF02288956](https://doi.org/10.1007/BF02288956).
- McCrae, R.R., 2015. A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review* 19, 97–112. doi:[10.1177/1088868314541857](https://doi.org/10.1177/1088868314541857).
- McNemar, Q., 1946. Opinion-attitude methodology. *Psychological Bulletin* 43, 289–374. doi:[10.1037/h0060985](https://doi.org/10.1037/h0060985).
- Mottus, R., Kandler, C., Bleidorn, W., Riemann, R., McCrae, R.R., 2017. Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology* 112, 474–490. doi:[10.1037/pspp0000100](https://doi.org/10.1037/pspp0000100).
- Möttus, R., Sinick, J., A.Terracciano, Hřebícková, M., Kandler, C., Jang, J.A....K.L., 2019. Personality characteristics below facets: A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability, and utility of personality nuances. *Journal of Person-*

- ality and Social Psychology 117. doi:[10.1037/pspp0000202](https://doi.org/10.1037/pspp0000202).
- Nagel, M., Watanabe, K., Stringer, S., Posthuma, D., Van Der Sluis, S., 2018. Item-level analyses reveal genetic heterogeneity in neuroticism. *Nature Communications* 9, 905–910. doi:[10.1038/s41467-018-03242-8](https://doi.org/10.1038/s41467-018-03242-8).
- R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- 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, 203 – 212. doi:[10.1016/j.jrp.2006.02.001](https://doi.org/10.1016/j.jrp.2006.02.001).
- Rentfrow, P.J., Gosling, S.D., Jokela, Markus; Stillwell, D.J., Kosinski, M., Potter, J., 2013. Divided we stand: Three psychological regions of the United States and their political, economic, social, and health correlates. *Journal of Personality and Social Psychology* 105, 996–1012. doi:[10.1037/a0034434](https://doi.org/10.1037/a0034434).
- Revelle, W., 2020a. psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston. <https://CRAN.r-project.org/package=psych>. R package version 2.0.1.
- Revelle, W., 2020b. psychTools: Tools to Accompany the Psych Package for Psychological Research. Northwestern University, Evanston. <https://CRAN.r-project.org/package=psychTools>. R package version 2.0.1.
- Revelle, W., Condon, D., Wilt, J., French, J.A., Brown, A.D., Elleman, L.G., 2016. Web and phone based data collection using planned missing designs, in: Nigel G. Fielding, Raymond M. Lee, G.B. (Ed.), *The Sage Handbook of Online Research Methods*. 2nd ed.. SAGE Publications. chapter 33, pp. 578–595.
- Revelle, W., Condon, D.M., 2019. Reliability from alpha to omega: A tutorial. *Psychological Assessment*. doi:[10.1037/pas0000754](https://doi.org/10.1037/pas0000754).
- Revelle, W., Wilt, J., Rosenthal, A., 2010. Individual differences in cognition: New methods for examining the personality-cognition link, in: Gruszka, A., Matthews, G., Szymura, B. (Eds.), *Handbook of Individual Differences in Cognition: Attention, Memory and Executive Control*. Springer, New York, N.Y.. chapter 2, pp. 27–49.
- Rhemtulla, M., Savalei, V., Little, T.D., 2016. On the asymptotic relative efficiency of planned missingness designs. *Psychometrika* 81, 60–89. doi:[10.1007/s11336-014-9422-0](https://doi.org/10.1007/s11336-014-9422-0).
- 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, 117–143.
- Spearman, C., 1904. “General Intelligence,” objectively determined and measured. *American Journal of Psychology* 15, 201–292. doi:[10.2307/1412107](https://doi.org/10.2307/1412107).
- Spearman, C., 1910. Correlation calculated from faulty data. *British Journal of Psychology* 3, 271–295. doi:[10.1111/j.2044-8295.1910.tb00206.x](https://doi.org/10.1111/j.2044-8295.1910.tb00206.x).
- Strong, E.K., 1927. Vocational interest test. *Educational Record* 8, 107–121.
- Thurstone, L.L., 1933. *The theory of multiple factors*. Edwards Brothers, Ann Arbor, Michigan.
- Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73, 273–282. doi:[10.1111/j.1467-9868.2011.00771.x](https://doi.org/10.1111/j.1467-9868.2011.00771.x).
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 301–320. doi:[10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x).

7. Appendix

Here we include the R code for our analyses. We make use of two R packages: *psych* (Revelle, 2020a) and *psychTools* (Revelle, 2020b).

The data from the spi data set are in *psychTools*, the bigger data set is available from DataVerse.

The code is a little redundant to be more readable.

7.1. Reliability analyses for the SPI data set

First find α and ω_h for each of the Big 5. Then find α for the little 27.

R code

```
library(psych) #make this package active
library(psychTools) #make this one active as well
$find alpha and omega for each scale
omA <- omega(spi[selectFromKeys(spi.keys$Agree)],plot=FALSE)
omC <- omega(spi[selectFromKeys(spi.keys$Consc)],plot=FALSE)
omN <- omega(spi[selectFromKeys(spi.keys$Neuro)],plot=FALSE)
omE <- omega(spi[selectFromKeys(spi.keys$Extra)],plot=FALSE)
omO <- omega(spi[selectFromKeys(spi.keys$Open)],plot=FALSE)
omega.h <- c(omA$omega.h,omC$omega.h,omN$omega.h,omE$omega.h,omO$omega.h)
omega.t <- c(omA$omega.tot,omC$omega.tot,omN$omega.tot,omE$omega.tot,
             omO$omega.tot)
alphas <- c(omA$alpha,omC$alpha,omN$alpha,omE$alpha,omO$alpha)
omega.df <- data.frame(omega_total = omega.t,alpha=alphas,omega_h = omega.h)
rownames(omega.df) <- cs(Agreeableness, Conscientiousness, Neuroticism,
                        Extraversion, Openness)
#find the scale scores for the Big 5 and little 27
spi.scales <- scoreItems(spi.keys,spi) #find scores as well as scale statistics
#combine demographics and scores
#The demographic information is in the first 10 columns of spi
spi.scores <- data.frame(spi[1:10],spi.scales$scores)

R5<- lowerCor(spi.scores[11:15])
basic.stats <- cbind(omega.df,R5)
```

7.2. Regression analyses for the SPI data set

Use the scales we found with `scoreItems` as predictors in the regressions. The criteria are the first 10 variables in the spi data set. We use `setCor` to find the multiple correlations and beta weights. We do not plot the results.

Then do the cross validation, taking the second random half of the spi data.

R code

```
spi.scales <- scoreItems(spi.keys,spi) #find scores as well as scale statistics
#combine demographics and scores

sc.demos <- data.frame(spi[1:10],spi.scales$scores)
#sc.demos <-cbind(spi[1:10],sc$scores) #combine with scores with demographics
```

```

set.seed(42) #for reproducible results
ss <- sample(1:nrow(sc.demos),nrow(sc.demos)/2)
#derivation multiple Rs
sc.5 <- setCor(y=1:10,x=11:15,, data=sc.demos[ss,], plot=FALSE)
sc.27 <- setCor(y=1:10,x=16:42, data=sc.demos[ss,], plot=FALSE)
sc.135 <- setCor(y=1:10,x=11:145,data=spi[ss,], plot=FALSE)

#now cross validate
cv.5 <- crossValidation(sc.5,sc.demos[-ss,])
cv.27 <- crossValidation(sc.27,sc.demos[-ss,])
cv.135 <- crossValidation(sc.135,spi[-ss,])
#combine them into one data frame
cross.valid.df <- data.frame(cv5=cv.5$crossV, cv.27=cv.27$crossV,
                             cv135=cv.135$crossV)

bs <- bestScales(spi[ss,],criteria=colnames(spi)[1:10], folds=10, n.item=20,
                 dictionary=spi.dictionary,cut=.05)
bs.cv <- crossValidation(bs,spi[-ss,])

#sort them in ascending order of the Big 5 multiple R
cross.valid.df.bs <- cbind(cross.valid.df,bs=bs.cv$crossV)
cv.df.bs.sorted <- dfOrder(cross.valid.df.bs,1)
matPlot(cv.df.bs.sorted[c(2,4,6,8)],minlength=8,
        main="Cross validation of multiple regression on spi data",
        xlas=3, ylab="Cross Validated R", pch=15:19)

legend(1,.6,cs(135,27,bestS,b5),lty=c(3,2,4,1),col=c(3,2,4,1))

bs.spi.smoke <- bs$items$smoke

df2latex(bs.spi.smoke[c(2,3,5)])

```

7.3. Graphical displays: the Manhattan plot

R code

```

labels <- names(spi.keys)
labels <- abbreviate(labels,minlength=8)
op <- par(mfrow=c(2,3)) #two row by three column display
man <- manhattan(spi,criteria=cs(health,exer,smoke),
                 keys=spi.keys,abs=FALSE,labels=labels)
man <- manhattan(spi,criteria=cs(health,exer,smoke),keys=spi.keys,abs=FALSE,
                 labels=labels,log.p = TRUE,main="")
op <- par(mfrow=c(1,1)) #put it back to the normal condition

```

7.4. Studies 3 and 4: Predicting 19 criteria from 696 items and scales

Now, do this for the 126K cases in the bigger sapa data set We get this by going to [Condon and Revelle \(2015\)](#), [Condon et al. \(2017a,b\)](#) and getting the 3 rda files there. We then stitch these three together using `rbind` to create the full.sapa data

We first need to get the data from *Dataverse* We download 3 Rdata files and the rbind them together

R code

```
#get the files from Dataverse and then combine them
full.sapa <- rbind(sapaTempData696items08dec2013thru26jul2014,
  sapaTempData696items26jul2014thru22dec2015,
  sapaTempData696items22dec2015thru07feb2017)

sapa <-char2numeric(full.sapa) #convert string character data to numeric data

spi.items <- selectFromKeys(spi.keys)
#set the number of multiple cores to take advantage of them
#does not work on PCs
options("mc.cores" = 4)
#find the scale scores and reliabililties
sapa.spi <- scoreItems(spi.keys, sapa, impute="none")
spi.scores <- sapa.spi$scores
demos <- sapa[c(2:10,14:23)] #choose the 19 demographic data
demos.spi <- cbind(demos, spi.scores)
set.seed(42)
ss <- sample(1:nrow(demos.spi), nrow(demos.spi)/2)
demos.b5 <- setCor(y=1:19,x=20:24,data=demos.spi[ss,],
  plot=FALSE) #do the multiple regressions
summary(demos.b5)
demos.27 <- setCor(y=1:19,x=25:51,data=demos.spi[ss,],plot=FALSE)

demos.135 <-setCor(y =1:19,x=spi.items,data=sapa[ss,],plot=FALSE)

#now cross validate
pred.b5 <- predict.psych(demos.b5,data=demos.spi[-ss,] )
pred.27 <- predict.psych(demos.27,data=demos.spi[-ss,] )
cross.valid.b5 <- diag(cor2(pred.b5,demos.spi[-ss,1:19]))
cross.valid.127 <- diag(cor2(pred.27,demos.spi[-ss,1:19]))
```

R code

```
sapa <- read.file() #goes to my directory to find the file
load(sapa) #one extra step required
sapa <- char2numeric(sapa) #makes the fields numeric
criteria <- colnames(sapa)[c(2:10,14:23)] #choose 19 criteria
spi.items <- selectFromKeys(spi.keys)
options("mc.cores"=8) #I am using a mac with multiple cores
scores <- scoreIrt.2pl(spi.keys,sapa) # Do IRT scoring of the data

big.scores <- cbind(sapa[criteria],scores)

set.seed(42) #for reproducible results
ss <- sample(1:nrow(big.scores), nrow(big.scores)/2)

#derivation multiple Rs
sc.5 <- setCor(y=criteria,x=20:24, data=big.scores[ss,], plot=FALSE)
sc.27 <- setCor(y=criteria,x=25:51, data=big.scores[ss,], plot=FALSE)
sc.135 <- setCor(y=criteria, x=spi.items,data=sapa[ss,],plot=FALSE)

#now cross validate
cv.5 <- crossValidation(sc.5,big.scores[-ss,])
cv.27 <- crossValidation(sc.27,big.scores[-ss,])
cv.135 <- crossValidation(sc.135,sapa[-ss,])
```

```

cross.valid.df <- data.frame(cv5=cv.5$crossV, cv.27=cv.27$crossV,
                             cv135=cv.135$crossV)
cross.valid.df.sorted <- dfOrder(cross.valid.df,1)

#ItemInfo is a dictionary of all of the sapa items
#this information is also on DataVerse
ItemInfo <- sapa.icar.dictionary
bs.sapa<- bestScales(sapa[ss,],criteria=criteria, folds=10, n.item=20,
                    dictionary=ItemInfo[,1:2],cut=.05)
bs.cv <- crossValidation(bs.sapa,sapa[-ss,])

#combine the best scales
cross.valid.df <- data.frame(cv5=cv.5$crossV, cv.27=cv.27$crossV,
                             cv135=cv.135$crossV,cvbs= bs.cv$crossV)
cross.valid.df.sorted <- dfOrder(cross.valid.df,1)
matPlot(cross.valid.df.sorted[c(2,4,6,8)],
        main="Cross validation of multiple regression on sapa data",
        xlas=3, ylab="Cross Validated R",pch=15:18)
legend(1, .5,cs(bestS,27,135,b5),lty=c(4,2,3,1),col=c(4,2,3,1),pch=c(18,16,17,15))
#now try profiles
R.big <- cor(sapa[ss,24:719],sapa[ss,criteria],use="pairwise")
R.pheno <- cor(sapa[ss,criteria],use="pairwise")
R.profile <- cor(R.big)
sapa.pheno.profile <- lowerUpper(R.pheno,R.profile)
corPlot(sapa.pheno.profile,xlas=3,
        main="Phenotypic (lower) and Profile (upper) correlations")

```

7.5. Study 5: Applying large sample profiles to a smaller sample

Now, lets try to do cross sample validities

R code

```

#we need to change sex to gender to the next step with the spi
colnames(spi)[2] <- "gender"
colnames(sc.demos)[2] <- "gender"

small.crit <- criteria[criteria %in% colnames(spi)[1:10]]
#find the profiles
R.spi.sapa <- cor(sapa[spi.items],sapa[small.crit],use="pairwise")

```

For the small criteria,find the Big5, little 27 and 135 regressions and best scales We can can do this for the full sample since we are applying to different sample

R code

```

small.5 <- setCor(y=small.crit,x = 20:24,data=big.scores,plot=FALSE)
small.27 <- setCor(y=small.crit,x = 25:51,data=big.scores,plot=FALSE)
small.135 <- setCor(y=small.crit,x=spi.items,data=sapa,plot=FALSE)
small.prof <- R.spi.sapa
small.bs <- bestScales(sapa[c(small.crit,spi.items)],criteria=small.crit,
                      folds=10, n.item=20,dictionary=ItemInfo[,1:2],cut=.05)

```

```

#now validate with the spi data
valid.5 <- crossValidation(small.5,sc.demos)
valid.27 <- crossValidation(small.27,sc.demos)
valid.135 <- crossValidation(small.135,spi)
valid.bs <- crossValidation(small.bs,spi)
small.prof.valid <- crossValidation(small.prof,spi)
small.valid.df <- data.frame(cv5=valid.5$crossV, cv.27=valid.27$crossV,
cv135=valid.135$crossV,cvbs= valid.bs$crossV,profile = small.prof.valid$crossV)
small.valid.df.sorted <- dfOrder(small.valid.df,1)

matPlot(small.valid.df.sorted[c(2,4,6,8,10)],
        main="Cross validation of multiple regression from SAPA on spi data",
        xlas=3, ylab="Cross Validated R",pch=15:19)
legend(1,.5,cs(bestS,profile,27,135,b5),lty=c(4,5,2,3,1),col=c(4,5,2,3,1),
      pch=c(18,19,16,17,15))

```