Web and phone based data collection using planned missing designs

William Revelle, David M. Condon, Joshua Wilt, Jason A. French, Ashley Brown and Lorien G. Elleman

The past few years have seen a revolution in the way that we are able to collect data. Using diaries (Bolger, Davis, & Rafaeli, 2003; Green, Rafaeli, Bolger, Shrout, & Reis, 2006) or smart phones (Mehl & Conner, 2012; Wilt, Funkhouser, & Revelle, 2011) to measure states within subjects across multiple time periods or the web to collect measures on thousands of subjects at a time (Gosling, Vazire, Srivastava, & John, 2004; Rentfrow, Gosling, & Potter, 2008; Revelle, Wilt, & Rosenthal, 2010; Wilt, Condon, & Revelle, 2011) has led to an exciting explosion in the amount of data collected. However, most of these studies ask the same questions of all of their participants.

In this chapter we review an alternative approach where we intentionally give each participant just a small subset of the items of interest but, with the power of basic psychometrics and sampling theory, are able to analyze the data as if far more items were presented. We refer to this procedure as Synthetic Aperture Personality Assessment (SAPA) (Condon & Revelle, 2014; Revelle et al., 2010) to emphasize the use of synthetic covariance matrices. That is, we find the correlations between composite scales, not based upon scoring the raw items, but rather by synthetically finding the covariances between scales based upon basic covariance algebra applied to the pair-wise complete item covariances. We think of these techniques as analogous to the techniques used in radio astronomy where the resolving power (aperture) of a set of radio telescopes may be greatly increased by synthesizing the signals collected by each individual telescope. Indeed, by combining the signals of radio telescopes scattered around the world, the effective aperture of these long baseline radio telescopes is the size of the entire earth. Because our covariance matrices are

based upon data sets with a great deal of intentionally missing data we also refer to our data as Massively Missing Completely at Random (MMCAR).

Our approach is not new, for it was discussed by Frederic Lord (1955) and then elaborated (Lord, 1977) in the assessment of ability. A variant of the technique that uses Balanced Incomplete Blocks (BIB) or "spiraling" has been applied in large scale international surveys such as the Programme for International Student Assessment (PISA) (Anderson, Lin, Treagust, Ross, & Yore, 2007). However, with the exception of our own work, we are not aware of the widespread use of this technique in smaller scale studies nor the complete emphasis on randomness that we have used. In this chapter we review the basic technique, discuss how to analyze the data, consider the effective sample size and resulting precision of estimates based upon scales and items, and then give a few examples of SAPA based results. We emphasize the application of these procedures to web based data collection because we have not yet implemented experience sampling or ecological momentary assessments more broadly defined with SAPA technique. However, we believe the techniques are relevant to both within-subject and between-subject means of data collection.

In the spirit of open science, all of the software we have developed, and all of the items we use, are in the public domain. We use open source software for data collection and analysis, and public domain items measuring temperament, ability and interests. In addition, we periodically publish the raw data to allow other researchers to use them (e.g., Condon & Revelle, 2015a,b,c, 2016).

Consider the basic problem of trying to determine the relationship between two or more constructs. In the past, psychological scales would be developed for each construct, the relevant items would be given to a relatively small set of subjects, and the covariances/correlations between these constructs would be found by scoring scales based upon the individual item responses. A typical procedure would include administering a number of inventories to a set of freshman in a group testing situation at the beginning of a school term. With the normal limitations of such a design, questionnaires could be given to a group of 100 - 500 students each of whom would answer all items given, probably at the rate of about 1-6 items per minute, depending upon their difficuty. The total testing time would limit the number of items given, and in an hour only several questionnaires, each with 20-40 items would be given. Another design, taking much longer, would be to recruit a community sample willing to take many questionnaires over the course of several years, e.g., the Eugene-Springfield sample (ESS) of Goldberg (1999). This procedure has led to a correlation matrix of several thousand items based upon nearly 800 subjects. A third technique, of course, is to use web based data collection from volunteers as in such studies as the German Socio-Economic Panel Study (Wagner,

Frick, & Schupp, 2007), or the http://www.outofservice.com/bigfive website which collects data for such studies as Rentfrow & Gosling (2003); Rentfrow et al. (2008) or the site run by John Johnson http://www.personal.psu.edu/faculty/j/5j/IPIP/ipipneo300.htm which presents either a 60 or 300 item version of the IPIP:NEO (Buchanan, Johnson, & Goldberg, 2005; Johnson, 2005). In all of these approaches, scales are found by combining scores on the individual items. But unfortunately, such volunteers are usually not willing to answer very many items and thus one is faced with a bandwidth versus fidelity tradeoff. One can either ask a few items each for many constructs, with the resulting low reliabilities, or many items for each of a few constructs with more reliability but less coverage.

Collecting MMCAR data using SAPA

An alternative procedure (SAPA) is to ask a few items for each construct from many subjects, but to randomly sample the items from a much larger pool of items. This allows for identification of the covariances between scales based on the composite covariances of the items rather than the raw item responses. This procedure takes advantage of the fact that people want to know about themselves (perhaps following the Delphic maxim to "know thyself") and makes up the lack of precision associated with giving few items with the abundance of traffic available on the web. Based upon the participant's responses, the SAPA website (sapa-project.org) offers customized and individualized personality feedback which was originally adapted from Buchanan et al. (2005) and Johnson (2005) but since greatly modified.

We do not actively advertise the site and have found that some of the traffic comes from people who have posted their feedback from us on their personal webpages, while others find it by searching the web for "personality tests" or "personality theory" etc. Recent evidence suggests that such self selected participants do not differ a great deal from those who are actively recruited to participate in probability-based national panel studies (Hays, Liu, & Kapteyn, 2015). For unfortunately, both means of data collection suffer from respondents willingness to participate and the reasons to opt into a sample are only slightly different from the reasons to opt out (Ansolabehere & Rivers, 2013). However, it is important to realize that not everyone is willing to participate in web based surveys (Pew Research Center, 2015). As would be expected, given that many of our participants are in college, the daily and monthly rates will vary during the year, but we have been averaging about 45,000 people participants a year.

A reasonable question is how valid are internet surveys in general, and ours in particular. We have compared our item structures and sample characteristics with those reported in the intensive study of the 800-1000 people who were individually

given many of the items we have used (the Eugene-Springfield sample of Goldberg & Saucier, 2016). As we discuss later, in terms of ethnicity, age, and education, our sample is much more diverse than the ESS, but the factor structures are remarkably similar. Additional validity data will come from as yet un-analyzed data of peer reports for a subset of our participants. Within the US, the distribution of our sample by state correlates with US Census population values .95. Our sample is certainly more diverse than is normally achieved at a selective research institution which tends to produce more WEIRD subjects (Henrich, Heine, & Norenzayan, 2010) than in our sample. However, it is certainly not representative of even the US population, for as we show below, our sample is younger, more educated, and has a higher proportion of females than does the US population.

The SAPA logic

Suppose one is interested in measuring facet level data from the "Big 5" measures of personality (the so called CANOE or OCEAN of personality – Conscientiousness, Agreeableness, Neuroticism, Openness and Extraversion – Digman, 1990; Goldberg, 1990) and eventually the relationship of these facets to measures of ability (Carroll, 1993; Gottfredson, 1997) and interests (Holland, 1959). Each facet might reflect 5-10 items, with two to five facets per broader domain, the measures of ability might include 50-100 items, and the measures of interests might involve 100-400 items. That is, the desired item pool is of the order of 400-600 items. But the typical subject is not willing to answer more than 40-75 items. The SAPA solution is to sample items completely at random from the larger pool (or perhaps systematically sample randomly from each of the temperament, ability and interest domains) and then present the items in random sets of 25 at a time. At the end of each set of 25 items subjects are asked if they want to continue, and if so, another 25 items are presented. They may stop whenever they want and feedback is presented to them based upon the items they have taken. Although the precision of measurement for each construct for each person is low, the precision of the synthetically formed covariances/correlations between scales measuring each construct is quite high.

How does this work? From the larger pool of P items, n items are selected with probability p_i, where $n = \bar{p}_i P = \sum_{i=1}^{P} (p_i)$, i.e., the average probability of any item being chosen, p_i, times the size of the total item pool. Thus, for N subjects filling out the questionnaire, each item has roughly $p_i N$ responses. More importantly, the average number of responses to each pair of items (i, j) is $p_i p_j N$. Consider the case of three months of data with $N = 10,000, P = 500$, and $p_i = p_j = .1$ or $n = 50$. Every one of the 500 items has been given roughly 1,000 $(p_i N)$ times and there are roughly 100 observations per pair of items $(p_i p_j N)$. (These numbers are given

merely for example purposes. In reality we tend to collect data for longer periods of time and build up about 500-1000 pairwise observations.) The subscript on the item probabilities reflect our relative interest in the content of the item. Demographic variables are presented with $p_i = 1$ while more exploratory items might be given with $p_i = .05$. When developing new ability items with a concern for their difficulty or when presenting items that are temporally relevant (e.g., attitudes towards an election), item presentation probabilities are increased and they might be presented with $p_i = .5$.

Item level statistics (e.g., the mean or variance) are based upon the $p_i N$ observations, while item inter-covariances are based upon $p_i p_j N$ pair-wise complete covariances. Structural analyses (e.g. factor analysis or principal components analysis) as well as internal consistencies of the individual scales (e.g., coefficients α and ω_h), and the correlations between individual scales may be found by basic matrix operations on the total inter-item covariance matrix rather than on the raw data matrix. This is not magic, but merely a function of covariance algebra.

In addition to the randomly chosen temperament, ability, and interest items, we also collect demographic information from all participants. These data include age, education, marital status, parental education, height, weight, smoking history, country and state of residence, and for those who say they are from the United States, their Zip Code. For these items, $p_i = 1$ and the precision of the resulting statistics are based upon the N participants measured.

Software used to present SAPA items

There are logically three different phases of presenting items and storing the individual responses. All three phases use open source software with specific code developed for this project. The phases are 1) specifying the item bank, 2) presenting the items and 3) storing the results and giving feedback.

Item bank

The item bank is stored using MySQL, an open source relational database management system which is supported by a large user community and has a commercial version as well. With the use of extensive help files from the MySQL community, programming is relatively easy. The data base is structured with a list of roughly 4,400 temperament, ability and interest items. 2,413 of the temperament items are taken from the open source International Personality Item Pool (IPIP, Goldberg, 1999).

The IPIP was developed by Lew Goldberg who adapted a short stem item format developed in the doctoral dissertation of Hendriks (1997) and items from the Five

Factor Personality Inventory developed in Groningen (Hendriks, Hofstee, & De Raad, 1999). Goldberg (1999) used about 750 items from the English version of the Groningen inventory, and has since supplemented them with many more new items in the same format. The initial development of the IPIP was controversial, as some believed that commercial developers could do a better job (Costa & McCrae, 1999). The citation count to the IPIP belies this belief. With at least 2,382 Google Scholar citations to the original publication (Goldberg, 1999) and 1,636 to the subsequent discussion (Goldberg et al., 2006) it is safe to say that open source personality measurement is a good idea. The IPIP items have been translated into at least 39 languages by at least 65 different research teams but the SAPA site is currently using just English based items. These were taken from `ipip.ori.org`.

We supplemented the IPIP item bank with 92 interest items taken from the Oregon Vocational Interest Scales (ORVIS, Pozzebon, Visser, Ashton, Lee, & Goldberg, 2010), 60 from the O*Net markers of Holland's RIASEC dimensions of interests (Armstrong, Allison, & Rounds, 2008; Holland, 1997; Rounds, Su, Lewis, & Rivkin, 2010), 60 music preference items (Liebert, 2006), 30 Right Wing Authoritarian items (Altemeyer, 2004), 78 items from the Eysenck Personality Questionnaire (Eysenck, Eysenck, & Barrett, 1985), 30 items taken from inventories (e.g., Jackson, 2009; Smederevac, Mitrović, Čolović, & Nikolašević, 2014) to measure aspects of Reinforcement Sensitivity Theory (Smillie, 2008), 220 items to measure aspects of personality disorders (Krueger, Derringer, Markon, Watson, & Skodol, 2013), 15 items from the Santa Barbara Sense of Direction scale (Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002), as well as 60 ability items developed as part of the International Cognitive Ability Resource project (ICAR, Condon & Revelle, 2014). Additional items that were taken from a number of different scales were given in prior years. The master list of the 4,300 items from the IPIP, ORVIS, O*Net and other sources that we use are available at `https://sapa-project.org/MasterItemList/`.

Presentation software

Using the server side scripting language, PHP: Hypertext Preprocessor, we query the MySQL server for items to present, and then display them using HTML5 on an APACHE based web server. Participant responses are then preprocessed and stored back to the MySQL server. As would be expected in any software development environment, our PHP scripts have improved over the years to take advantage of changes in MySQ, PHP, and to the Hyper Text Markup Language (HTML5). The site was originally hosted at the `personality-project.org` website and has

since been migrated to the sapa-project.org website. (Both of these are hosted at Northwestern University).

From the user's perspective, they see a number of screens with "radio button" response options, or a few text box options. These screens or "pages" include:

Welcome: A brief description of the SAPA project, a FAQ about the test, the research behind SAPA, links to literature about current research in individual differences and the benefits that may accrue to the user.

Consent form: A brief discussion about how long the test will take, how all responses are anonymous, that participants will receive feedback based upon our norms, and a consent button to start the test.

Demographics: One question is whether people have taken the survey before, others ask age (in a text box). Pull down menu options ask about gender, height, weight, marital status, relationship status, frequency of exercise, smoking history, country and state/region where the person grew up, level of education, university major (if relevant), employment status, general field of work, and then parental education. More recently we have started asking about the participants postal or ZIP code. At this point the user is assigned (invisibly) a random identification number (RID) which will be used to check for repeated entries in the same web browser session.

First and subsequent page of questions: Each page has 25 questions, the first 21 of which are sampled from the temperament and interest item banks, the final four of which are ability items sampled from our ability item bank. At the end of each of the first three pages, subjects are told that they will have more accurate feedback if they continue. At the end of the fourth page, they are given personality feedback based upon scores calculated from the items they have answered.

Optional subsequent pages: Participants are offered the possibility of continuing on and filling out more items about such things as creative accomplishments, or of sending a message to a friend to rate them on various personality attributes.

Storage and feedback

As the participant is filling out the survey, results are transmitted to the MySQL server at the end of every page and stored with their random identification number (RID). Once the participant selects the option saying that they are finished with the entire set of (randomly administered) items to which they chose to respond, they

are given scores on various personality scales. These were originally based upon the Big Five factors but have more recently been replaced with hierarchically–organized factors scores with three, six and 12 factors. This scoring is done by applying a key of all possible items for each scale and finding the average response given to the items that were presented. The graphic output gives a location of each of the scores on a line along with a confidence interval for each score.

Data security

When we first started the site and for the subsequent eight years, the SAPA project was hosted on an Apple MacIntosh desktop computer in the Personality, Motivation, and Cognition laboratory at Northwestern. We updated our security settings on APACHE, MySQL, and PHP relatively frequently, but not enough to prevent a MySQL injection from taking over the system. After recovering the data (with one weeks' worth lost to the hacker), we moved the site to a more professionally managed server at the main computer cluster on campus. We mention this as a warning of the problems of maintaining web servers.

Analyzing SAPA/MMCAR data

The basic logic of the SAPA procedure follows from some fundamental principles of psychometrics with respect to correlations of items and correlations of item composites. It is well known that the correlation between two scales, A and B with n and m items respectively, is $\frac{Cov_{AB}}{\sqrt{V_A V_B}}$. But since the covariance of two item composites is merely the sum of the covariances of the separate items, $Cov_{AB} = \sum_{j=1}^{n} \sum_{k=1}^{m} (cov_{a_i b_j})$ and, similarly, the variance of a composite is the sum of the variances and covariances of the items in that composite $Var_A = \sum_{j=1}^{n} \sum_{k=1}^{n} (cov_{a_i a_j})$, then

$$r_{AB} = \frac{Cov_{AB}}{\sqrt{V_A V_B}} = \frac{\sum_{j=1}^{n} \sum_{k=1}^{m} (cov_{a_i b_j})}{\sqrt{\sum_{j=1}^{n} \sum_{k=1}^{n} (cov_{a_i a_j}) \sum_{j=1}^{m} \sum_{k=1}^{m} (cov_{b_i b_j})}}. \tag{1}$$

More compactly, in matrix algebra, and for the general case of multiple scales, let the raw data be the matrix $\mathbf{X}$ with N observations on P items converted to deviation scores (with most rows having only n non-missing items). Then the item variance–covariance matrix is $\mathbf{C} = \mathbf{X}\mathbf{X}' N^{-1}$ and scale scores, $\mathbf{S}$ are found by $\mathbf{S} = \mathbf{K}'\mathbf{X}$. $\mathbf{K}$ is a keying matrix, with $\mathbf{K}_{ij} = 1$ if $item_i$ is to be scored in the positive direction for scale j, 0 if it is not to be scored, and -1 if it is to be scored in the negative direction. In this case, the covariance between scales, $\mathbf{C}_s$, may be found by pre and post multiplying the item covariance matrix with a matrix of the keys:

$$\mathbf{C}_s = \mathbf{K}'\mathbf{X}(\mathbf{K}'\mathbf{X})' N^{-1} = \mathbf{K}'\mathbf{X}\mathbf{X}'\mathbf{K} N^{-1} = \mathbf{K}'\mathbf{C}\mathbf{K}. \tag{2}$$

The scale correlations, R_s are found by pre and post multiplying the scale covariance matrix $\mathbf{C}_s$ by the inverse of the scale standard deviations, which are merely the square roots of the diagonal of $\mathbf{C}_s$:

$$R_s = (diag(\mathbf{C}_s))^{-.5}\mathbf{C}_s(diag(\mathbf{C}_s))^{-.5} \tag{3}$$

That is, the covariance between any set of scales can be found by multiplying the transposed keying matrix by the inter-item covariance matrix times the keying matrix. The correlations are found by dividing this product by the standard deviations.

Although the correlational structure of the *items* requires the raw data, the correlations of *scales* can be found by keying the item correlation matrix into scale correlations, not the raw data matrix. In the case of a SAPA/MMCAR design, this is very important, for while the individual item correlations can be found by "pairwise complete correlations" or "available case correlations", it is highly unlikely that any one participant has complete data for any scale. We conduct our structural analyses at the item and scale covariance level, rather than at the raw data level. We believe that the greater resolution of item-level and scale-level covariances made possible by our technique compensates for the lack of complete subject data analysis.

In order to process our SAPA data, we have developed a number of functions included in the *psych* package (Revelle, 2015) in the open source statistical system R (R Core Team, 2015). These functions are specifically meant to handle the massively missing data structures that we use and are referenced below. In addition, we have developed an additional package, *SAPA-tools* (French & Condon, 2015), to facilitate data extraction from the MySQL server and doing some basic data cleaning. Unless explicitly mentioned, the R functions discussed in the following pages are all from the *psych* package.

Data cleaning

After importing the data from the MySQL server into R, either using functions in the *RMySQL* (Ooms, James, DebRoy, Wickham, & Horner, 2015) package, the *SAPA-tools* package, or just reading the file using a normal HTML browser and copying to the clipboard, the data need some preliminary data checking and cleaning. Some participants will take the questionnaire, receive their feedback, and then go back to the beginning of the page to do it again. This is detected by keeping the RID permanent for the web browser session. Thus, the data are first cleaned by removing all duplicate RID numbers. (The data are, however, maintained so that we could, if we desire, go back and find out the characteristics of those who enter more than one set of questions.) Additional data cleaning procedures includes removing subjects

who report ages less than 14 or more than 90 and excludes those participants who tell us they have previously participated in the survey.

Basic item information

Descriptive statistics (means, standard deviations, ranges, etc.) are found for all items using the `describe` function. Demographic information is available for all participants, whereas temperament, ability, and interest items are given to just random subsets of participants. Pairwise counts of the frequency of particular item pairs are examined to facilitate further analysis. (Given the changing nature of items being administered, not all item subsets are administered together. This is particularly the case when doing exploratory studies.) Correlations between ability items are found using tetrachoric correlations; correlations between temperament and interest items are found by polychoric correlations. Correlations of continuous variables (e.g., age, height, weight) with dichotomous (ability) or polytomous (temperament and interests) items are found using polyserial correlations. All of these correlations are done using the `mixedCor` function.

Scale level structures

The real power of the SAPA procedure is evident when we examine the correlational structure either at the item or at the scale level. Factor analyses of the item level covariances are done using the `fa` function and two-parameter item response theory statistics based upon these factor analyses (McDonald, 1999) are done using the `irt.fa` function. For instance, the tetrachoric correlation matrix of dichotomous ability items may be factored using a minimum residual factor analysis function `fa` and the resulting loadings, λ_i are transformed to item discriminations by $a = \frac{\lambda}{\sqrt{(1-\lambda^2)}}$. The difficulty parameter, δ, is found from the τ parameter of the tetrachoric function and the factor loadings of the factor analysis of the tetrachoric matrix: $\delta = \frac{\tau}{\sqrt{(1-\lambda^2)}}$. Similar analyses may be done with polytomous item responses using polychoric correlations and distinct estimates of item difficulty (location) for each item response.

Similarly, analysis of internal structure of each scale may be done based upon the correlation matrices using functions to find α (`alpha`, `scoreItems`), $\omega_{hierarchical}$ and ω_{total} (`omega`)(Revelle & Zinbarg, 2009) as well as the signal/noise ratio of each scale (`scoreItems`). The hierarchical cluster structure based upon the item correlations (Revelle, 1979) is found using the `iclust` function. When examining the correlations of nested scales, that is scales with overlapping items because they might be subscales of other scales, we use a correction derived from Cureton (1966) and Bashaw & Anderson Jr (1967) (`scoreOverlap`).

Individual and group level scores

When describing the personality characteristics of certain subgroups (e.g., college majors, occupations, Zip codes), it is necessary to use scores based upon the raw data. To do this, we use IRT based estimates from the available items for each subject using `irt.fa` and `score.irt`. This procedure, although highly correlated with just adding up the item responses, allows slightly more precision in that it takes into account item discriminations and item endorsement frequencies (difficulties).

It is important to realize that the correlations between scales using the synthetic procedures may differ from those based upon the simple sum or IRT based scores. This is because of the missingness in the data. The individual level scores for a particular measure might be based upon 2-4 items, and the subsequent correlation with another similar scale, will be attenuated by the missingness in the data. However, the structural correlations, based upon the covariance of all of the items in the scale (as many as 20-50) will be much less attenuated.

Because of the sample size, it is also possible to find the correlational structure of the mean scores for groups organized by e.g., college major or occupation. These correlations are *between group* correlations and will not necessarily be the same, and indeed usually are not the same, as the correlations pooled *within group* or the *overall* correlations. Although some dismiss these correlations of aggregates as showing "the ecological fallacy" (Robinson, 1950) or the Yule-Simpson "paradox" (Simpson, 1951; Kievit, Frankenhuis, Waldorp, & Borsboom, 2013; Yule, 1903), we find that they tell us meaningful information about how individuals aggregate into groups (Revelle & Condon, 2015).

Precision of SAPA/MMCAR data

The standard error of the correlation between two particular items will be the classical standard error $\sigma_r = \frac{\sqrt{1-r^2}}{\sqrt{N-2}}$. For complete data, this is the same formula for the correlation of composite scales. But what about the standard errors of SAPA based composite scales? What is the appropriate sample size? Is it the number of participants who take any individual pair of items $(p_i p_j N)$ or is it somehow closer to N? To answer this question, we rely on simulation. The following is based partly on the work of Brown (2014) who has done a much more thorough simulation than is reported here.

For a population covariance matrix of 0 between two sets of items that correlate .3 within and 0 between, we took 1,000 random samples of 10,000 cases for complete data, and for data with a probability of observing a particular item of .1, .125, .25,

.5 and 1. That is, for the .1 condition, the probability of any pair of items having data was .01.

In addition, we simulated scales with 1, 2, 4, 8, or 16 items. Each of the 1,000 random samples governed by a particular combination of scale size and proportion of observed items produced a sample correlation calculated in one of two different ways: either as pair-wise complete, or using the full information maximum likelihood (FIML) method. Each sample scale correlation was also corrected for alpha reliability, and minres oblimin factor analyses sought a two-factor solution whenever scale size was 16. Four sets of statistics (uncorrected and corrected correlations, factor loadings and intercorrelations) and their standard errors were computed by taking the mean and standard deviation, respectively, of the appropriate set of 1,000 sample statistics.

Results indicated that uncorrected correlations derived using the SAPA method approach their latent values as scale size increases; that is, as one aggregates over more items. This suggests that analysts who do not correct for reliability would do well to aggregate over items as SAPA does. In addition, both uncorrected and corrected correlations' standard errors decrease as scale size increases; this effect seems to be more pronounced with larger quantities of missing data. In essence, aggregating over items increases effective sample size more than might be expected based solely on the number of cases and the probability of observing a given item (Figure 1). Effective sample size, N_e is merely a function of the standard error, σ_r, of the correlation, r, which is $\sigma_r = \sqrt{\frac{1-r^2}{N-2}}$. Thus $N_e = \frac{1-r^2}{\sigma_r^2} + 2$. We find the standard error by simulation to estimate the effective sample size. Finally, and as expected, more missing data tends to produce slightly more biased, less precise results among corrected correlations and factor intercorrelations. Factor loadings were less precise when more data were missing, but the effect of missing data on bias was, in this case, relatively small.

Also of interest here is the fact that the FIML method did not greatly improve upon the quality of the relevant statistics. Both statistical bias and data patterns, as described above, were the same regardless of analytic method. FIML produced slightly more precise solutions than the standard SAPA method, but it is much more computationally intense and time-consuming and, moreover, it is better-suited to the analysis of data that possesses only a few distinct patterns of missingness, as in the commonly-used balanced incomplete block design. Based upon our simulations (Brown, 2014), we propose that our method represents a simple and economical way for survey researchers with sample sizes of at least 200 to increase breadth of coverage without sacrificing statistical rigor. Obviously, for smaller sample sizes, the sampling probability for each item needs to be larger than we are using for our larger samples.
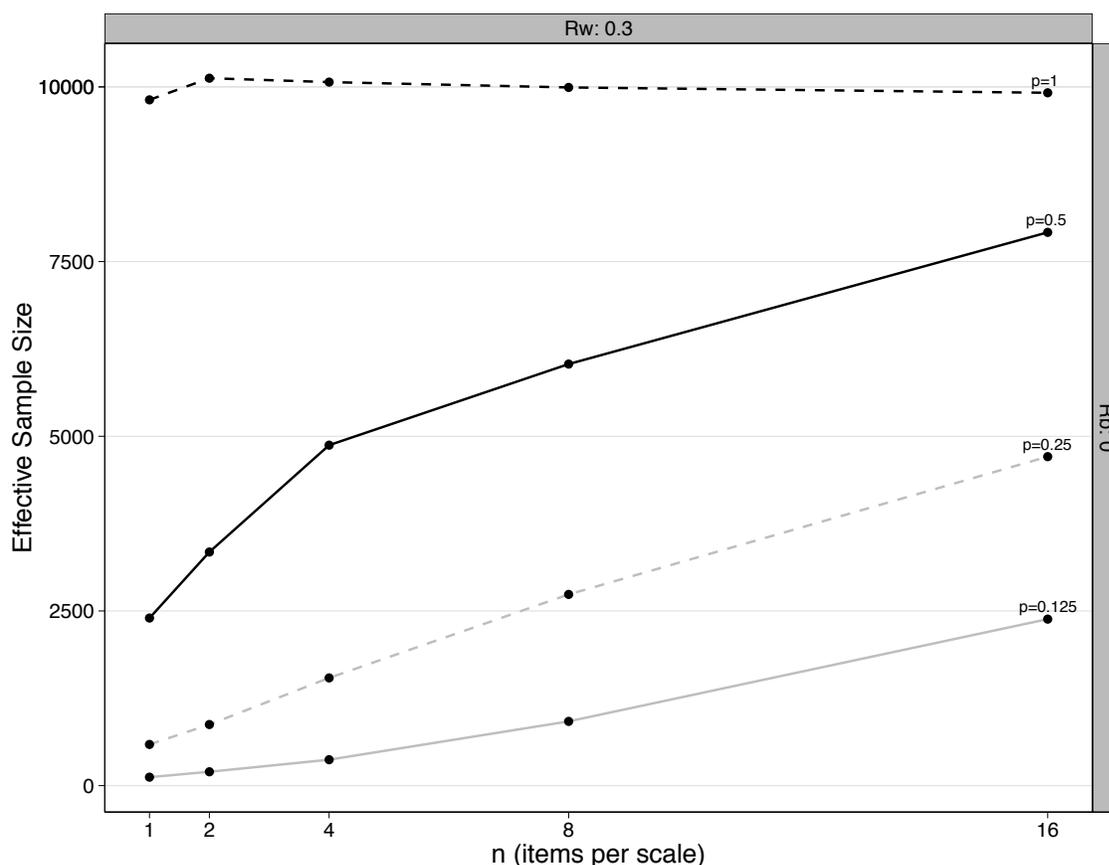
Figure 1. The effective sample size is a function of the observed standard error of the correlation and is $N_e = \frac{1-r^2}{\sigma_r^2} + 2$. For this particular simulation, the average within scale correlation was set to .3 and the average between set correlation to 0. Means shown are from 1,000 replications. What is important to observe is when using the MMCAR composite scales, that effective sample size increases dramatically as the number of items per scale is increased.

Examples of SAPA results

The following are short summaries of some the major projects conducted using SAPA. These include analysis of the correlates of items differing in their saturation of of affective, behavioral, cognitive and desire content (Wilt, 2014), and examinations of alternative structures of items administered in several different personality inventories (Condon, 2014). We have already reported the development of an open source ability test used in the SAPA project (Condon & Revelle, 2014) and are now using SAPA procedures to validate other item types. In addition, one of the powers

of the technique is that side studies can be conducted by introducing items with relatively low probabilities of being included and then just waiting a long time, or alternatively give some items with a high probability of being administered and then run them for just a few weeks.

Demographics of the SAPA participants

The demographics in this section are based on a sample of 207,002 participants, whose self-report data were collected between August 2010 and December 2015. Participants from this sample are 63% female. Participants grew up in 215 countries, with the United States accounting for 73% of the sample. Twenty-two countries besides the U.S. have 500 or more participants, with the top three being Canada (8,895), the United Kingdom (5,577), and Australia (4,024) . Participants from the U.S. identify as 67% white, 10% African American, 9% Hispanic, 5% Asian American, 1% Native Alaskan/Hawaiian/American, 6% multiracial, and 1% "other". The mean age of participants is 26 ($sd = 11$; $median = 22$). The age distribution is highly skewed for both males and females (Figure 2). The modal participant is between 19 and 22 and is currently in college or university (Table 1). This distribution is roughly the same for males and females (Figure 3).

These results highlight both a strength and weakness of voluntary web based data collection. In terms of age and gender it is clearly the case that our data are not representative of the population. However, it is also the case that our data represent much greater diversity of subject characteristics than found in the typical university based sample or even community based samples such as the ESS.

**Table 1.** Highest education attained, by age

Education Level	[14-18]	[19-22]	[23-29]	[30-39]	[40-49]	[50-90]
< 12 years	25,319	463	325	236	179	165
High School graduate	7,363	2,477	1,715	1,152	715	624
Currently in University	13,263	44,210	16,747	9,016	4,279	1,870
Some College	164	2,166	3,202	2,547	1,538	1,205
College Degree	152	3,086	9,810	6,599	3,683	2,571
In Graduate Or Professional School	164	1,089	4,860	2,031	887	386
Graduate Or Professional Degree	42	271	3,570	5,019	3,322	3,009

Participants' Age by Gender
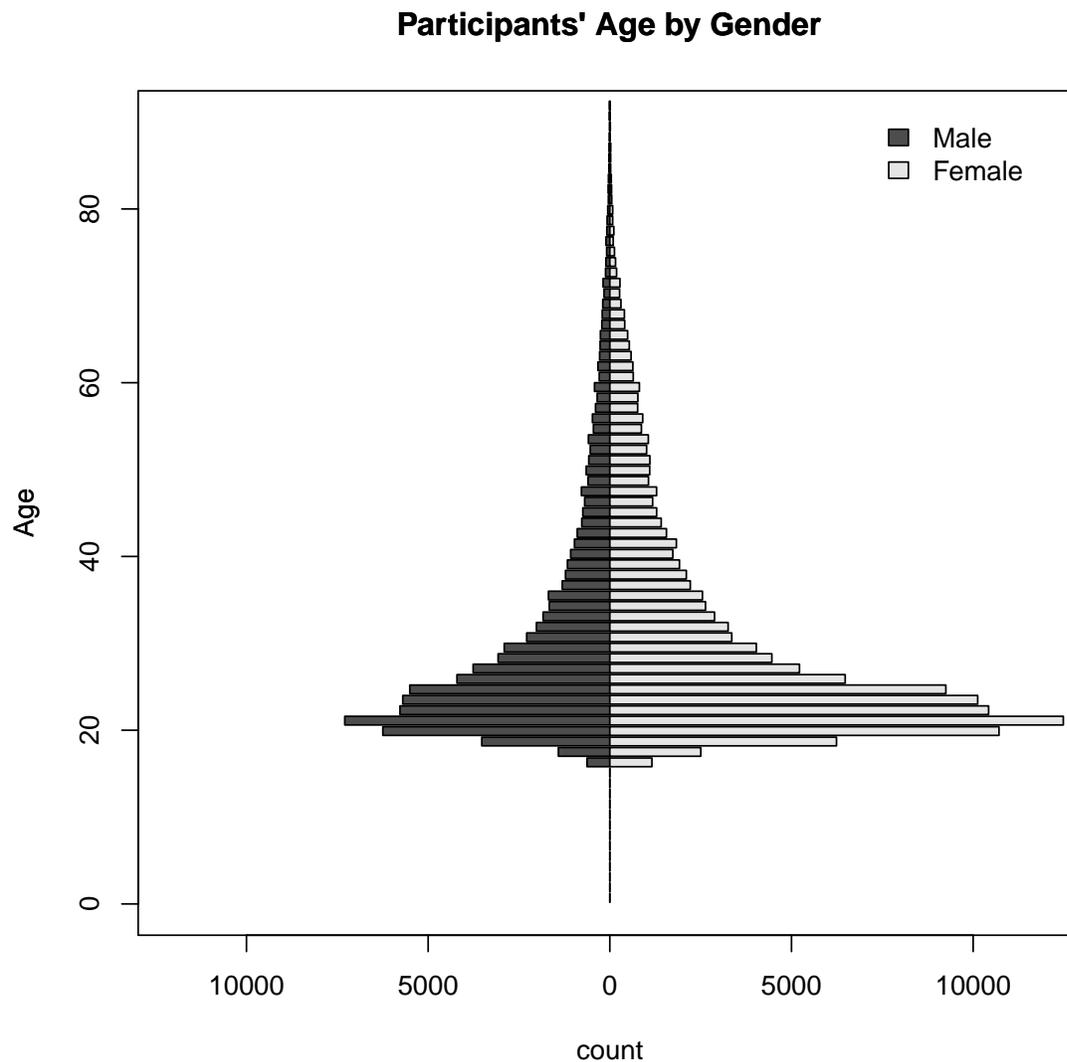
Figure 2. Although there are roughly twice as many females as males, the age age distributions are roughly the same,

Personality Questionnaires and the ABCDs

Personality traits have been conceptualized as individual differences in patterns of affect (A), behavior (B), cognition (C) and desire (D) over time and space

Participants' Education by Gender
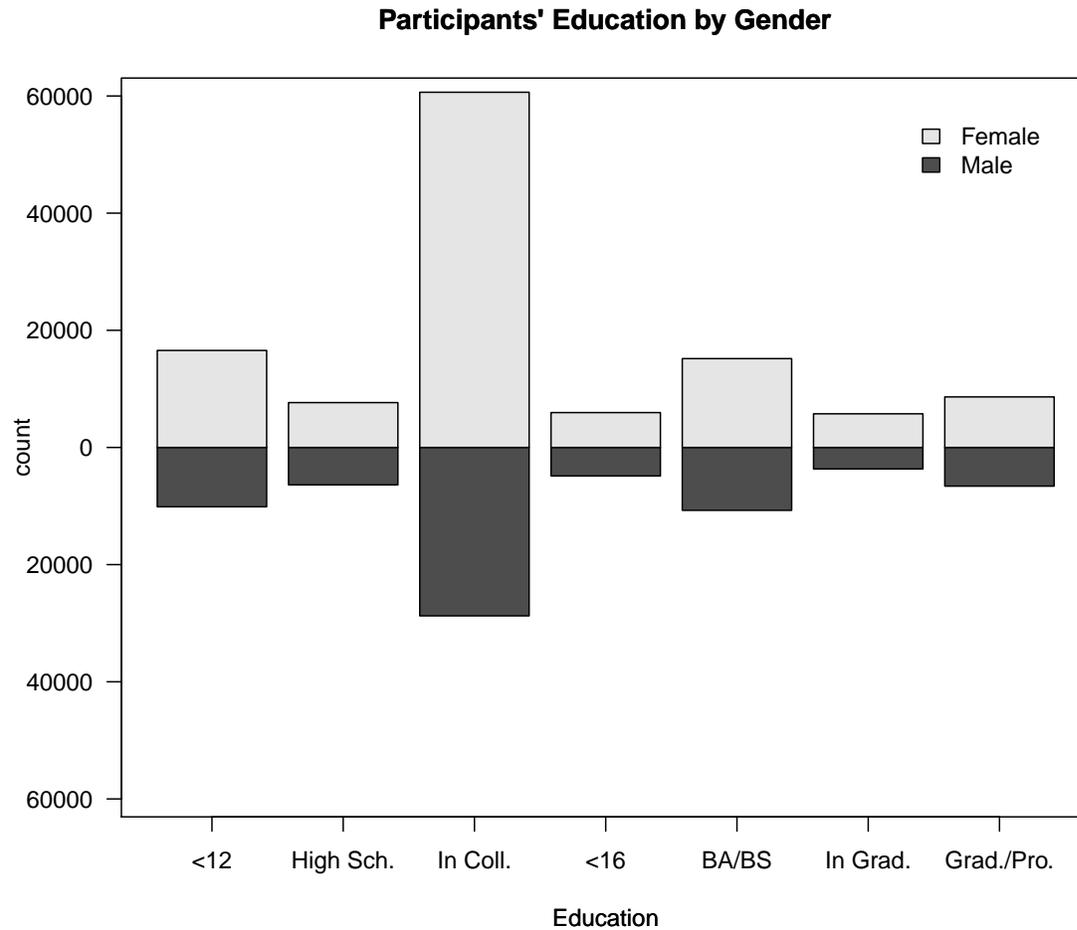
Figure 3. The female-to-male ratio of participants is highest for college students.

(Allport, 1937; Johnson, 1997; Winter, John, Stewart, Klohnen, & Duncan, 1998; Revelle, 2008), yet the most common assessments of the Big-Five traits (Costa & McCrae, 1992b; Goldberg, 1992) do not explicitly refer to these ABCD components (Pytlik Zillig, Hemenover, & Dienstbier, 2002). We therefore conducted a content analysis in order to identify items for each Big Five trait that reflected primarily one A, B, C, or D content (Wilt, 2014; Wilt & Revelle, 2015). We identified items from each ABCD domain for each trait and created facet scales from these items: for example, the ABCD facet scales of agreeableness were labeled as sympathetic

affect, considerate *behavior*, trusting *cognition*, and *desire*. Using the *psych* package (Revelle, 2015) in R (R Core Team, 2015), we employed the SAPA technique to generate a synthetic correlation matrix containing the ABCD items assessing the Big Five. From this correlation matrix, we determined that (i) a Big Five structure emerged from factor analysis of the items; (ii) even when correcting for item overlap, using the `scoreOverlap` function, Big Five trait domain scales correlated highly with their respective ABCD facet scales, (iii) ABCD scales within each trait were positively correlated with each other, and (iv) items had strong correlations with their respective ABCD facet scale.

The factor structure of personality inventories

A primary goal of the SAPA-project for the past several years has been to examine the structural properties of a number of personality inventories which share overlapping items in the IPIP. Athough there are 1,034 items contained in eight different inventories: 100 in the Big 5 marker scales (Goldberg, 1999), 100 in the Big 5 aspect scales (DeYoung, Quilty, & Peterson, 2007), 240 in the IPIP-HEXACO (Ashton, Lee, & Goldberg, 2007), 300 in the IPIP-NEO (Goldberg, 1999), 127 in the IPIP-Multidimensional Personality Questionnaire (Tellegen & Waller, 2008), 48 items in the Questionnaire Big 6 scales (Thalmayer, Saucier, & Eigenhuis, 2011), and 79 in the Eysenck Personality Questionnaire (Eysenck et al., 1985); there are only 696 unique items. For instance, the 100 items in the Big Five factor markers (Goldberg, 1999) are the same as the ones used in the Big Five Aspects Scales (DeYoung et al., 2007). This set of 696 items includes all the items from at least 255 of the personality scales listed at the IPIP website, including IPIP items designed to match the above inventories as well as e.g.,, the Hogan Personality Inventory (Hogan & Hogan, 1995).

Based upon the correlation matrix of nearly 24,000 participants, we tested for the number of factors that would best represent the structure. Unfortunately, the exploratory factor structure of these 696 items did not yield any clean solution for the number of factors but the most interpretable solutions represented three, five or 15 factors (Condon, 2014). Most importantly, these solutions were not nested in the standard hierarchical representation reported by many, but were best described as forming a heterarchy. With careful item analysis, a set of 150 items was found to represent all levels of this heterarchy quite well, with scales that could reliably distinguish these three, five or 15 dimensions. The data for the 23,681 participants and the 696 items are available for others to use through *DataVerse*, an open source data repository (Condon & Revelle, 2015c,b).

Current work is underway to examine how these dimensions relate to differences in interests and desires across college majors and across occupational groups as these

appetites are reflected in the interest dimensions known as the RIASEC (Holland, 1997).

The International Cognitive Ability Resource

Cognitive ability assessment differs from many other measures of individual differences because it requires tools that evaluate maximal performance levels rather than levels of typical behavior. In other words, cognitive ability items are objectively scored as correct or incorrect. This important difference makes it more challenging to measure cognitive ability than other constructs; more test security is required to maintain fairness and validity. Still, researchers and clinicians are strongly motivated to employ cognitive ability measures that can provide quick, reliable and cost efficient assessment by virtue of the fact that the many aspects of cognitive ability are highly predictive of a wide range of outcomes (Deary, 2009).

Several brief electronic measures of cognitive ability are available, but few (if any) are both widely-validated and available for use in non-proctored environments (i.e., over the internet). None exist in the public-domain. The International Cognitive Ability Resource ("ICAR", Condon & Revelle, 2014) was developed in order to fill this absence. Following the initial creation of four item types that were validated using the SAPA-Project, an international collaboration of German, British and American universities has since been formed to encourage the development of a range of cognitive ability measures (see the `icar-project.com` for more information).

A prior report of the preliminary results was based upon the first 65,000 subjects (Revelle et al., 2010) and a subsequent report discussed the validation of the expanded inventory with another 24,000 participants (Condon & Revelle, 2014). The data from that article are available in an open source respository (Condon & Revelle, 2015a, 2016). Sample data from this project are also available as the `ability` data set in the *psych* package.

Examples of Scale Development and Validation Across Broad Domains

The breadth of constructs that can be simultaneously assessed using the the SAPA methodology allows for evaluation of the relative contribution of factors across broad domains of individual differences. By domains, we allude to the affective, cognitive and conative domains which have long been recognized in the social sciences (McDougall, 1923; Condon, 2014; Holland, 1959, 1997; Carroll, 1993; McGrew, 2009; Costa & McCrae, 1992a; Digman, 1990; Goldberg, 1990). If not for the use of the SAPA technique, cross-sectional evaluation of the contribution of these broad domains of individual differences to achievement would be impractical as thorough evaluation of cognitive abilities, vocational interests and temperament in addition

to achievement outcomes would require several hours of participants' time and attention. We have previously reported SAPA studies on music preferences (Liebert, 2006; Revelle et al., 2010) and trust (Evans & Revelle, 2008) and have since extended these to studies of psychopathy (Wright, 2014) and creative achievements (Zaonegina, Condon, Zabelina, & Revelle, 2015).

Summary and conclusions

Telemetric methods have revolutionized the ways in which we can collect data. However, there is a natural tendency to continue our traditional reluctance to have missing observations even as we collect orders of magnitude more data. We believe that this is a mistake and have outlined the power of using a massively missing, completely at random item administration technique. We have shown the power of introducing such missingness into our designs. We have emphasized the methodology and results from our web based project (SAPA) but believe that similar techniques would be useful with modern smart phone apps.

Acknowledgements

Funding

References

Allport, G. W. (1937). *Personality; a psychological interpretation.* New York,: H. Holt and Company.

Altemeyer, B. (2004). Highly dominating, highly authoritarian personalities. *Journal of Social Psychology*, *144*(4), 421-447. doi: 10.3200/SOCP.144.4.421-448

Anderson, J., Lin, H., Treagust, D., Ross, S., & Yore, L. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education*, *5*(4), 591–614. doi: 10.1007/s10763-007-9090-y

Ansolabehere, S., & Rivers, D. (2013). Cooperative survey research. *Annual Review of Political Science*, *16*(1), 307-329. doi: 10.1146/annurev-polisci-022811-160625

Armstrong, P. I., Allison, W., & Rounds, J. (2008). Development and initial validation of brief public domain RIASEC marker scales. *Journal of Vocational Behavior*, *73*(2), 287 - 299. doi: 10.1016/j.jvb.2008.06.003

Ashton, M. C., Lee, K., & Goldberg, L. R. (2007). The IPIP-HEXACO scales: An alternative, public-domain measure of the personality constructs in the HEXACO model. *Personality and Individual Differences*, *42*(8), 1515-1526. doi: 10.1016/j.paid.2006.10.027

Bashaw, W., & Anderson Jr, H. E. (1967). A correction for replicated error in correlation coefficients. *Psychometrika*, *32*(4), 435–441. doi: 10.1007/BF02289657

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579-616. doi: 10.1146/annurev.psych.54.101601.145030

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

Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the internet. *European Journal of Psychological Assessment*, *21*(2), 115-127. doi: 10.1027/1015-5759.21.2.115

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY, US: Cambridge University Press. doi: 10.1017/CBO9780511571312

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

Condon, D. M., & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52-64. doi: 10.1016/j.intell.2014.01.004

Condon, D. M., & Revelle, W. (2015a). Selected ICAR data from the SAPA-Project: Development and initial validation of a public-domain measure. *Harvard Dataverse*. doi: 10.7910/DVN/AD9RVY

Condon, D. M., & Revelle, W. (2015b). Selected personality data from the SAPA-Project: 08dec2013 to 26jul2014. *Harvard Dataverse*. doi: 10.7910/DVN/SD7SVE

Condon, D. M., & Revelle, W. (2015c, August). Selected personality data from the SAPA-Project: On the structure of phrased self-report items. *Journal of Open Psychology Data*, *3*(1). doi: 10.5334/jopd.al

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*. doi: 10.5334/jopd.25

Costa, P. T., & McCrae, R. R. (1992a). Four ways five factors are basic. *Personality and Individual Differences*, *13*(6), 653-665. doi: 10.1016/0191-8869(92)90236-I

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

Costa, P. T., & McCrae, R. R. (1999). Reply to Goldberg. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, p. 29-31). Tilburg, The Netherlands: Tilburg University Press.

Cureton, E. (1966). Corrected item-test correlations. *Psychometrika*, *31*(1), 93–96. doi: 10.1007/BF02289461

Deary, I. J. (2009). Introduction to the special issue on cognitive epidemiology. *Intelligence*, *37*, 517–519. doi: 10.1016/j.intell.2009.05.001

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *Journal of Personality and Social Psychology*, *93*(5), 880-896. doi: 10.1037/0022-3514.93.5.880

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

Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, *42*(6), 1585-1593. doi: 10.1016/j.jrp.2008.07.011

Eysenck, S. B., Eysenck, H. J., & Barrett, P. (1985). A revised version of the Psychoticism scale. *Personality and Individual Differences*, *6*(1), 21-29. doi: 10.1016/0191-8869(85)90026-1

French, J. A., & Condon, D. M. (2015). *SAPA Tools: Tools to analyze the SAPA Project. R package version 0.1.*

Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216-1229. doi: 10.1037/0022-3514.59.6.1216

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, *4*(1), 26-42. doi: 10.1037/1040-3590.4.1.26

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, p. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84-96. doi: 10.1016/j.jrp.2005.08.007

Goldberg, L. R., & Saucier, G. (2016, January). *The Eugene-Springfield Community Sample: Information Available from the Research Participants* (Tech. Rep. No. 56-1). Eugene, Oregon: Oregon Research Institute.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, *59*(2), 93–104. doi: 10.1037/0003-066X.59.2.93

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79 - 132. doi: DOI:10.1016/S0160-2896(97)90014-3

Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, *11*(1), 87-105. doi: 10.1037/1082-989X.11.1.87

Hays, R. D., Liu, H., & Kapteyn, A. (2015). Use of internet panels to conduct surveys. *Behavior Research Methods*, *47*(3), 685-690. doi: 10.3758/s13428-015-0617-9

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, *30*(5), 425–447. doi: 10.1016/S0160-2896(02)00116-2

Hendriks, A. A. J. (1997). *The construction of the five-factor personality inventory (FFPI).* Unpublished doctoral dissertation, Rijksunivsiteit Groningen, Groningen, The Netherlands.

Hendriks, A. A. J., Hofstee, W. K., & De Raad, B. (1999). The five-factor personality inventory (FFPI). *Personality and Individual Differences*, *27*(2), 307 - 325. doi: 10.1016/S0191-8869(98)00245-1

Henrich, J., Heine, S. J., & Norenzayan, A. (2010, 6). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. doi: 10.1017/S0140525X0999152X

Hogan, R., & Hogan, J. (1995). *The Hogan personality inventory manual (2nd. ed.).* Tulsa, OK: Hogan Assessment Systems.

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology*, *6*(1), 35 - 45. doi: 10.1037/h0040767

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments.* Psychological Assessment Resources.

Jackson, C. J. (2009). Jackson-5 scales of revised reinforcement sensitivity theory (r-rst) and their application to dysfunctional real world outcomes. *Journal of Research in Personality*, *43*(4), 556 - 569. doi: 10.1016/j.jrp.2009.02.007

Johnson, J. A. (1997). Units of analysis for the description and explanation of personality. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (p. 73-93). San Diego, CA: Academic Press.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103-129. doi: 10.1016/j.jrp.2004.09.009

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*, *4*(513), 1-14. doi: 10.3389/fpsyg.2013.00513

Krueger, R., Derringer, J., Markon, K., Watson, D., & Skodol, A. (2013). *The personality inventory for DSM-5 (PID-5).* Washington, DC: APA.

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

Lord, F. M. (1955). Sampling fluctuations resulting from the sampling of test items. *Psychometrika*, *20*(1), 1–22. doi: 10.1007/BF02288956

Lord, F. M. (1977). Some item analysis and test theory for a system of computer-assisted test construction for individualized instruction. *Applied Psychological Measurement*, *1*(3), 447–455. doi: 10.1177/014662167700100313

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, N.J.: L. Erlbaum Associates.

McDougall, W. (1923). *Outline of psychology.* Oxford: Scribners.

McGrew, K. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1-10. doi: 10.1037/h0040767

Mehl, M. R., & Conner, T. S. (2012). *Handbook of research methods for studying daily life.* New York: Guilford Press.

Ooms, J., James, D., DebRoy, S., Wickham, H., & Horner, J. (2015). RMySQL: Database Interface and MySQL Driver for R [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=RMySQL (R package version 0.10.3)

Pew Research Center. (2015, September). *Coverage error in internet surveys.* Retrieved from http://www.pewresearch.org/files/2015/09/2015-09-22_coverage-error-in-internet-surveys.pdf

Pozzebon, J. A., Visser, B. A., Ashton, M. C., Lee, K., & Goldberg, L. R. (2010). Psychometric characteristics of a public-domain self-report measure of vocational interests: The oregon vocational interest scales. *Journal of Personality Assessment*, *92*(2), 168-174. doi: 10.1080/00223890903510431

Pytlik Zillig, L. M., Hemenover, S. H., & Dienstbier, R. A. (2002). What do we assess when we assess a big 5 trait? A content analysis of the affective, behavioral and cognitive processes represented in the big 5 personality inventories. *Personality and Social Psychology Bulletin*, *28*(6), 847-858. doi: 10.1177/0146167202289013

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*(6), 1236-1256. doi: 10.1037/0022-3514.84.6.1236

Rentfrow, P. J., Gosling, S. D., & Potter, J. (2008). A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics. *Perspectives on Psychological Science*, *3*(5), 339-369. doi: 10.1111/j.1745-6924 .2008.00084.x

Revelle, W. (1979). Hierarchical cluster-analysis and the internal structure of tests. *Multivariate Behavioral Research*, *14*(1), 57-74. doi: 10.1207/s15327906mbr1401\ _4

Revelle, W. (2008). The contribution of reinforcement sensitivity theory to personality theory. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality* (p. 508-527). Cambridge: Cambridge University Press.

Revelle, W. (2015). psych: Procedures for personality and psychological research [Computer software manual]. http://cran.r-project.org/web/packages/psych/. (R package version 1.5.8)

Revelle, W., & Condon, D. M. (2015). A model for personality at three levels. *Journal of Research in Personality*, *56*, 70-81. doi: 10.1016/j.jrp.2014.12.006

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* (p. 27-49). New York, N.Y.: Springer.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega and the glb: comments on Sijtsma. *Psychometrika*, *74*(1), 145-154. doi: 10.1007/ s11336-008-9102-z

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351–357.

Rounds, J., Su, R., Lewis, P., & Rivkin, D. (2010). *O* NET® interest profiler short form psychometric characteristics: Summary.* Raleigh, NC: National Center for O* NET Development.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *13*(2), 238–241.

Smederevac, S., Mitrović, D., Čolović, P., & Nikolašević, Ž. (2014). Validation of the measure of revised reinforcement sensitivity theory constructs. *Journal of Individual Differences*, *35*(1), 12-21. doi: 10.1027/1614-0001/a000121

Smillie, L. D. (2008). What is reinforcement sensitivity? neuroscience paradigms for approach-avoidance process theories of personality. *European Journal of Personality*, *22*(5), 359-384. doi: 10.1002/per.674

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the multidimensional personality questionnaire. *The Sage handbook of personality theory and assessment*, *2*, 261–292.

Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of brief to medium-length Big Five and Big Six personality questionnaires. *Psychological Assessment*, *23*(4), 995–1009. doi: 10.1037/a0024165

Wagner, G. G., Frick, J. R., & Schupp, J. (2007). *The German Socio-Economic Panel Study (SOEP) - evolution, scope and enhancements* (Tech. Rep.). The German Socio-Economic Panel Study DIW Berlin 10108 Berlin, Germany: SOEPpaper No. 1. doi: 10.2139/ssrn.1028709

Wilt, J. (2014). *A new form and function for personality.* Unpublished doctoral dissertation, Northwestern University.

Wilt, J., Condon, D., & Revelle, W. (2011). Telemetrics and online data collection: Collecting data at a distance. In B. Laursen, T. D. Little, & N. Card (Eds.), *Handbook of developmental research methods* (p. 163-180). New York: Guilford Press.

Wilt, J., Funkhouser, K., & Revelle, W. (2011). The dynamic relationships of affective synchrony to perceptions of situations. *Journal of Research in Personality*, *45*, 309–321. doi: 10.1016/j.jrp.2011.03.005

Wilt, J., & Revelle, W. (2015). Affect, behavior, cognition and desire in the Big 5: An analysis of item content and structure. *European Journal of Personality*, *29*(4), 478-497. doi: 10.1002/per.2002

Winter, D. G., John, O. P., Stewart, A. J., Klohnen, E. C., & Duncan, L. E. (1998). Traits and motives: Toward an integration of two traditions in personality research. *Psychological Review*, *105*(2), 230-250. doi: 10.1037/0033-295X.105.2.230

Wright, Z. E. (2014, May). *Creating a self-report measure of psychopathy using items from the Personality Inventory for the DSM-5.* Honors Thesis. Department of Psychology, Northwestern University.

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika*, *2*(2), 121–134.

Zaonegina, E., Condon, D. M., Zabelina, D., & Revelle, W. (2015). The creative achievement inventory and individual differences: Associations across and within the domains of creativity. *submitted*.