Web and phone based data collection using planned missing designs

Journal Title
XX(X):1–19

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

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 et al., 2003; Green et al., 2006) or smart phones (Mehl and Conner, 2012; Wilt et al., 2011b) to measure states within subjects across multiple time periods or the web to collect measures on thousands of subjects at a time (Gosling et al., 2004; Rentfrow et al., 2008; Revelle et al., 2010; Wilt et al., 2011a) 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 will 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 and 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. 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 et al., 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 the subsequent pages 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 an app to apply SAPA techniques to smart phones. However, we believe the techniques are relevant to both within-subject (e.g., smart phones) and between-subject (e.g., the web) 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 for use by other researchers (e.g., Condon and Revelle, 2015a,0).

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 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 et al.), or the `http://www.outofservice.com/bigfive` website which collects data for such studies as Rentfrow and Gosling (2003); Rentfrow et al. (2008) or the site run by John Johnson `http://www.personal.psu`

.edu/faculty/j/5/j5j/IPIP/ipipneo300.htm which presents either a 60 or 300 item version of the IPIP:NEO (Buchanan et al., 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 he covariances between scales based on he 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.~~ As would be expected, the daily and monthly rates will vary during the year, but we have been averaging about 45,000 people participants a year.

Suppose one is interested in measuring facet level data from the "Big 5" measures of personality and eventually the relationship of these facets to measures of ability and interests. 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 $p = 1$ while more exploratory items might be given with p = .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 = .5.

Internal consistencies of the individual scales, and the correlations between individual scales may be found by basic 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 which we will discuss later.

In addition to the randomly chosen temperament, ability, and interest items, we also collect demographic information from all participants. These data include age, education, 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, $p_i = 1$.

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 5,000 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 et al., 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 and McCrae, 1999). The citation count to the IPIP belies this belief. With at least 2141 Google Scholar citations to the original publication (Goldberg, 1999) and 1430 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 et al., 2010) as well as 60 ability items developed as part of the International Cognitive Ability Resource project (ICAR, Condon and Revelle, 2014).

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 evironment, our PHP scripts have improved over the years to take advantage of changes in MySQL and to the Hyper Text Markup Language (HTML). 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. 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 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 have responded are scored on three, six and twelve personality scales. 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 feedback was originally adapted from that of~~ Johnson (2005) ~~but has since been modified to emphasize a quasi-hierarchical structure.~~

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 $Var_a = \sum_{j=1}^{n} \sum_{k=1}^{n} (cov_{a_i a_j})$

$$r_{ab} = \frac{Cov_{ab}}{\sqrt{V_a V_b}} = \frac{\sum_{j=1}^{n} \sum_{k=1}^{m} (cov_{a_i b_j})}{\sqrt{\sum_{j=1}^{n} \sum_{k=1}^{n} (cov_{a_i a_j}) \sum_{j=1}^{m} \sum_{k=1}^{m} (cov_{b_i b_j})}}. \quad (1)$$

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 converted to deviation scores. 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, $\mathbf{C}_s$, is

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

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

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

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.

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. In addition, we have developed an additional package, *SAPA-tools* (French and Condon, 2015), to facilitate data extraction from the MySQL server and doing some basic data cleaning.

Data cleaning

After importing the data from the MySQL server into R, either using functions in the *RMySQL* (Ooms et al., 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.) Additonal 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 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. The tetrachoric correlation matrix of dichotomous 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 or polychoric function and the factor loadings of the tetrachoric matrix: $\delta = \frac{\tau}{\sqrt{(1-\lambda^2)}}$. Similar analyses may be done with discrete 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 and 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 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 and 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. 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 et al., 2013; Yule, 1903), we find that they tell us meaningful information about how individuals aggregate into groups (Revelle and Condon, 2015).

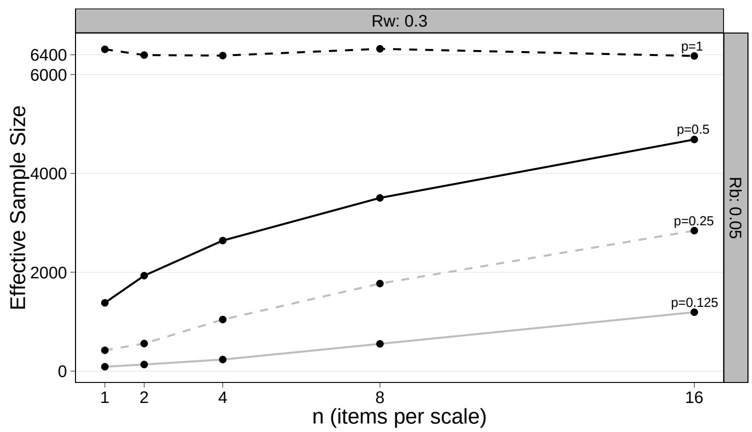
Precision of SAPA/MMCAR data

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 500 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 described above, or using the full information maximum likelihood (FIML) method. Each sample 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 500 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 alpha 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 (see Figure XX). Finally, and as expected, more missing data tends to produce 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.



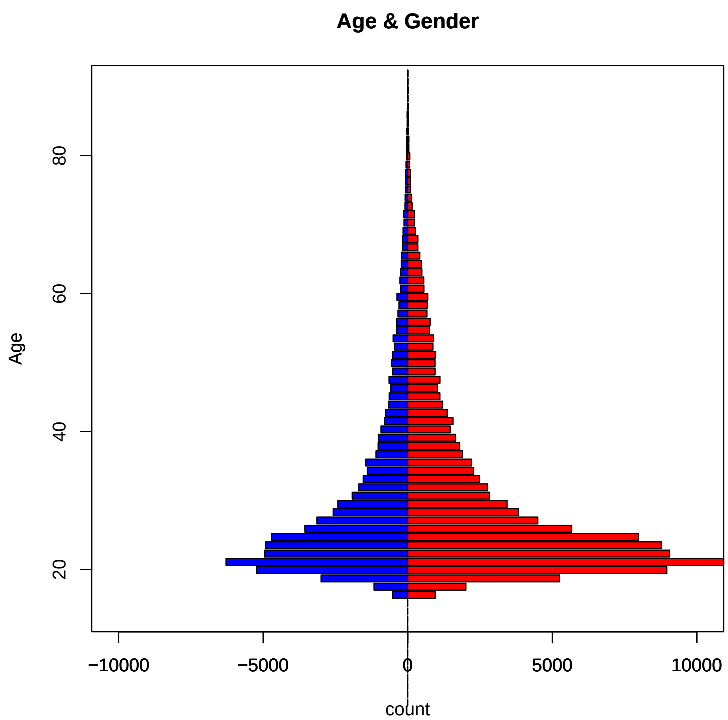
Also of interest here is the fact that the FIML method did not greatly improve upon the quality of the relevant statistics. Both statistical bias and data patterns, as described above, were the same regardless of analytic method. FIML produced slightly more precise solutions than the standard SAPA method, 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. We propose that our method represents a simple and economical way for survey researchers with sample sizes of at least 1000 to increase breadth of coverage without sacrificing statistical rigor.

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), 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 and Revelle, 2014) and will discuss the possibilities of 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 177,048 participants, whose self-report data were collected between August 2010 and April 2015. Participants from this sample are 63% female. Participants grew up in 214 countries, with the United States accounting for 74% of the sample. Sixteen countries besides the U.S. have 500 or more participants, with the top three being Canada (7,703), the United Kingdom (4,561), and Australia (3,338) . Participants from the U.S. identify as 67% white, 10% African American, 9% Hispanic, 6% multiracial, and 4% Asian American. The mean age of participants is 26 ($sd = 11$). Figure (XX) shows a histogram of gender by age, with blue bars representing males and red bars representing females. Table (XX) shows highest education attained by age. Table (XX) shows highest education attained by gender.



	[14,18]	(18,22]	(22,29]	(29,39]	(39,49]	(49,90]
less12yrs	21806	407	270	202	163	148
HSgrad	6333	2098	1429	993	618	540
CurrentInUniv	11481	38735	14683	8069	3809	1661
SomeCollege	141	1892	2706	2194	1348	1039
CollegeDegree	128	2648	8275	5624	3192	2187
InGradOrProSchool	136	952	4183	1769	772	324
GradOrProDegree	36	240	3059	4263	2823	2559

	male	female
less12yrs	8779	14217
HSgrad	5490	6521
CurrentInUniv	25053	53385
SomeCollege	4200	5120
CollegeDegree	9035	13019
InGradOrProSchool	3157	4979
GradOrProDegree	5600	7380

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 (Allport, 1937; Johnson, 1997; Winter et al., 1998; Revelle, 2008), yet the most common assessments of the Big-Five traits (Costa and McCrae, 1992; Goldberg, 1992) do not explicitly refer to these ABCD components (Pytlik Zillig et al., 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 and Revelle, 2015). We identified 7 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. These findings together suggest that the ABCD scales measured the Big Five and their respective ABCD content with good fidelity. In response to confusion concerning the key ingredients of personality traits (Yang et al., 2014), these ABCD scales clearly define the dominant psychological contents of items measuring each Big Five trait.

Integrating the affective, cognitive, and conative domains

david

Condon (2014)

Validating an ability inventory

david

Condon and Revelle (2014)

Side studies

One of the powers of the SAPA design is the ability to focus on the relationships of particular sets of items to the broader temperament, ability and interest domains.

Trust Evans and Revelle (2008)

Creativity

Psychopathology Wright (2014)

Music preferences Liebert (2006) music

Summary and conclusions

We have outlined the power of using a massively missing, completely at random item administration technique. We have emphasized out web based project (SAPA) but believe that simliar technqes would be useful with modern smart phone apps.

References

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

Anderson, J., Lin, H., Treagust, D., Ross, S., and 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.

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

Bolger, N., Davis, A., and Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology Vol 54 2003, 579-616.*

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

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

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

Condon, D. M. and Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, 43:52–64.

Condon, D. M. and Revelle, W. (2015a). Selected personality data from the SAPA-Project: 08dec2013 to 26jul2014. *Harvard Dataverse*.

Condon, D. M. and Revelle, W. (2015b). Selected personality data from the SAPA-Project: On the structure of phrased self-report items. *Journal of Open Psychology Data*.

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

Costa, P. T. and McCrae, R. R. (1999). Reply to Goldberg. In Mervielde, I., Deary, I., De Fruyt, F., and Ostendorf, F., editors, *Personality psychology in Europe*, volume 7, pages 29–31. Tilburg University Press, Tilburg, The Netherlands.

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

Evans, A. M. and Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, 42(6):1585–1593.

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

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

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., and Ostendorf, F., editors, *Personality psychology in Europe*, volume 7, pages 7–28. Tilburg University Press, Tilburg, The Netherlands.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., and 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.

Gosling, S. D., Vazire, S., Srivastava, S., and 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.

Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., and Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, 11(1):87–105.

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

Hendriks, A. A. J., Hofstee, W. K., and De Raad, B. (1999). The five-factor personality inventory (FFPI). *Personality and Individual Differences*, 27(2):307 – 325.

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

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

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., and Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*, 4(513):1–14.

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

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

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.

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

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

Ooms, J., James, D., DebRoy, S., Wickham, H., and Horner, J. (2015). *RMySQL: Database Interface and MySQL Driver for R*. R package version 0.10.3.

Pozzebon, J. A., Visser, B. A., Ashton, M. C., Lee, K., and 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.

Pytlik Zillig, L. M., Hemenover, S. H., and 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.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rentfrow, P. J. and 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.

Rentfrow, P. J., Gosling, S. D., and 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.

Revelle, W. (1979). Hierarchical cluster-analysis and the internal structure of tests. *Multivariate Behavioral Research*, 14(1):57–74.

Revelle, W. (2008). The contribution of reinforcement sensitivity theory to personality theory. In Corr, P. J., editor, *The Reinforcement Sensitivity Theory of Personality*, chapter 18, pages 508–527. Cambridge University Press, Cambridge.

Revelle, W. (2015). *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, http://cran.r-project.org/web/packages/psych/. R package version 1.5.6.

Revelle, W. and Condon, D. M. (2015). A model for personality at three levels. *Journal of Research in Personality*, 56:70–81.

Revelle, W., Wilt, J., and Rosenthal, A. (2010). Individual differences in cognition: New methods for examining the personality-cognition link. In Gruszka, A., Matthews, G., and Szymura, B., editors, *Handbook of Individual Differences in Cognition: Attention, Memory and Executive Control*, chapter 2, pages 27–49. Springer, New York, N.Y.

Revelle, W. and Zinbarg, R. E. (2009). Coefficients alpha, beta, omega and the glb: comments on Sijtsma. *Psychometrika*, 74(1):145–154.

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

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

Wagner, G. G., Frick, J. R., and Schupp, J. The German Socio-Economic Panel Study (SOEP) - evolution, scope and enhancements. Technical report, The German Socio-Economic Panel Study DIW Berlin 10108 Berlin, Germany.

Wilt, J. (2014). *A new form and function for personality.* PhD thesis, Northwestern University.

Wilt, J., Condon, D., and Revelle, W. (2011a). Telemetrics and online data collection: Collecting data at a distance. In Laursen, B., Little, T. D., and Card, N., editors, *Handbook of Developmental Research Methods*, chapter 10, pages 163–180. Guilford Press, New York.

Wilt, J., Funkhouser, K., and Revelle, W. (2011b). The dynamic relationships of affective synchrony to perceptions of situations. *Journal of Research in Personality*, 45:309–321.

Wilt, J. and Revelle, W. (2015). Affect, behavior, cognition and desire in the Big 5: An analysis of item content and structure. *European Journal of Personality*, (in press).

Winter, D. G., John, O. P., Stewart, A. J., Klohnen, E. C., and Duncan, L. E. (1998). Traits and motives: Toward an integration of two traditions in personality research. *Psychological Review*, 105(2):230–250.

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

Yang, Y., Read, S. J., Denson, T. F., Xu, Y., Zhang, J., and Pedersen, W. C. (2014). The key ingredients of personality traits: Situations, behaviors, and explanations. *Personality and Social Psychology Bulletin*, 40(1):79–91.

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