


## Personality Psychology

# The Convergence of Self and Informant Reports in a Large Online Sample

Anne Zola<sup>1</sup> <sup>a</sup>, David M. Condon<sup>2</sup>, William Revelle<sup>1</sup>

<sup>1</sup> Northwestern University, Evanston, IL, US, <sup>2</sup> University of Oregon, Eugene, OR, US

Keywords: online assessment, self-reports, informant-reports

<https://doi.org/10.1525/collabra.25983>

---

## Collabra: Psychology

Vol. 7, Issue 1, 2021

---

Despite their added benefits, informant-reports are largely underutilized in personality research. We demonstrate the feasibility of collecting informant-reports online, where researchers have unprecedented access to large, global populations. Using an entirely free, opt-in procedure tied to an existing personality survey, we collected 1,554 informant-reports for 921 unique targets, in conjunction with over 158,000 self-reports. Informant-reports showed a strong correspondence to self-reported traits at three levels of analysis: among the Big Five domains, the lower-level SPI-27 factors (Condon, 2018), and at the item-scale level. Among the Big Five, self-informant agreement ranged between .63 and .72, except for Openness (.42). Higher informant-ratings of Extraversion were positively associated with all Big Five self-ratings in the direction of social desirability. Across the Big Five and the 27 lower-order traits, agreement was strongest between self-reports of compassion and informant-reports of agreeableness (.74) and weakest between self-reported emotional expressiveness and informant-reported emotional stability (.02). Agreement between informants was roughly equivalent for all of the Big Five traits (.29 to .35) and attractiveness (.37), though agreement between informants for perceived intelligence was non-significant. In addition, we empirically identified the self-report items that best predict what informants say about targets, highlighting the features of self-reported personality that are most readily confirmed by informants. Finally, we discuss group level differences of participants who interacted with the informant-report system at various levels. In general, participants who sought and provided informant reports are more open and agreeable than the general sample, though targets' personality did not affect whether or not invited informants provided ratings.

Self-reports are the most widely used method in personality psychology, largely due to their relative ease and cost-efficiency. Informant-reports, by contrast, are underutilized in personality research despite their numerous benefits. Many researchers falsely assume informant-reports are substantially more difficult to collect than self-reports, even in the context of increasingly easy data collection online. Here, we demonstrate the feasibility of collecting informant-reports from a large, global sample using an entirely opt-in procedure and establish the convergent validity of these informant-reports with self-report data at three levels of analysis.

Multiple models of personality emphasize an outside observer's ability to provide non-redundant, relevant information about a target's traits (McAbee & Connelly, 2016; Vazire, 2010). Most build on the idea that the individual's internal perspective (self-report) provides insight on one's identity, whereas an outside perspective (informant-report) offers a look at one's reputation (Hogan, 2007). For some

purposes, this outside perspective may be an even better assessment of an individual's traits. In one meta-analytic review (Connelly & Ones, 2010), informant-reports surpassed self-reports as predictors for academic achievement and work performance.

Informant-reports have a long-standing reputation as being difficult to collect, though it is unclear whether this reputation is deserved. Vazire (2006) debunked multiple preconceptions about the time, effort, and cost required to collect informant-reports, and several prior studies demonstrate the feasibility of collecting moderate to large samples of informant-reports (Clifton et al., 2005; K. Lee & Ashton, 2017; Vazire, 2006). Yet, the impact on subsequent personality research has been negligible. In Vazire's (2006) analysis of the studies published in the *Journal of Research in Personality* (JRP) in 2003, only 24% of studies included informant-reports. The situation has not improved over the last 15 years. Among the 2017 JRP studies that collected data, only 11 (16%) used informant-reports; 65 (98%) used

---

a [anne.zola@northwestern.edu](mailto:anne.zola@northwestern.edu)

self-reports and, among these, 55 (83%) collected *only* self-reports. Studies published in the Personality Processes and Individual Differences section of the *Journal of Personality and Social Psychology* show a similar trend. Among the studies that collected data in 2018, only 4 (13%) used informant-reports; 30 (94%) used self-reports and 22 of these relied solely on self-reports. One possible explanation is that the aforementioned examples made of prior work using informant reports do not generalize well; perhaps the historical reliance on undergraduate samples, for example, differs from online data collection practices in ways that are relevant for informant-reports. Or, perhaps the evidence that informant-reports are under-utilized in personality research was not sufficiently widely recognized to overcome the perception that they are difficult to implement.

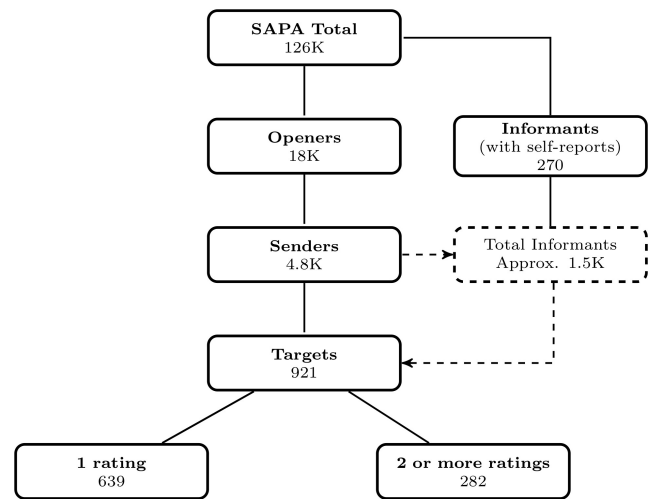
How can the situation be improved? One solution is to establish informant-report procedures that eliminate the need for increased costs or time. Reducing added burdens to near zero would make large-scale informant report collection more feasible. Moreover, it would help to show that large informant-report samples need not be restricted to in-person samples (typically involving undergraduate students). With over 5 billion internet users worldwide (Internet World Stats, 2021), over half the world's population is available to researchers via the web, for little to no cost. Online, researchers can offer participants the chance to “opt-in” to an informant-report system and send rating requests to friends, family, colleagues, and others. As part of the Synthetic Aperture Personality Assessment (SAPA) Project (<https://sapa-project.org>), we collected 1,554 informant-reports for 921 unique targets using this completely opt-in, web-based procedure. Our aim is two-fold: to demonstrate the feasibility of collecting large samples of informant-reports from an online, global population using a convenience sampling technique and to show the congruence of self-informant reports at three levels of analysis: at the Big-Five, among the 27 domains of a lower-order assessment model, and at the item level. In addition, we seek to describe group-level differences between the participants who interact with the opt-in informant-report system and the general sample.

## Method

### Procedure

The self-report and informant-report data described here were collected between May 20, 2013 and February 6, 2017. Participants were motivated to visit the site in order to learn more about their personality, and were provided with customized feedback upon completion of the self-report survey. All participation was free and anonymous. After completing the survey but before receiving feedback about their personality based on their responses, participants were asked to consider sending an email to acquaintances, friends and/or family members asking them to complete a brief questionnaire about the participant's personality. After informed consent, participants who completed a self-report or informant-report were assigned a random identification number (RID).

Unlike many other online surveys, the SAPA Project is designed with the intention that no one participant will complete all available items (Revelle et al., 2016). Instead,



**Figure 1. Nested subgroups with sample size.**

Arrows indicate selection (e.g., senders select total informants, total informants select which senders become targets).

SAPA participants receive only a random subset of the total items available. During this period of data collection, participants were given a random sample of 696 public domain items, mainly from the International Personality Item Pool (IPIP; Goldberg, 1999), and 60 cognitive ability items from the International Cognitive Ability Resource (ICAR; Condon & Revelle, 2014). Participants were required to respond to at least 22 items in order to receive feedback, but were allowed to complete as many as 300 personality items in total, plus as many as 26 optional demographic items. The subsequent data are characterized as massively missing, as no one participant completed the entire set of available items. However, as each participant was presented with a random subset of the possible items, the resulting missingness is distributed completely at random. Because the level of missingness is greater than 80-90%, we refer to our data as Massively Missing Completely at Random (MMCAR). Unbiased correlation/covariance matrices at the item level were constructed from these MMCAR data, which can be scored in order to find the correlation of the constructs.

The opt-in nature of the informant-report system created various subsets of participants (Figure 1): those who opened the request page (“Openers”), those who sent rating requests (“Senders”), those who were rated by others (“Targets”), and those who chose to rate one or more targets (“Informants”). After completing an informant-report, a number of informants also completed the SAPA self-report survey. Participants were organized in nested subsets, wherein Senders are members of the group Openers, and Targets are members of the group Senders. This research complied with current APA standards of ethical treatment and was approved (determined exempt) by Northwestern University's Institutional Review Board. See supplementary materials to access the R script, which includes code for downloading the data directly into the R environment through an API.

### Measures

Informants were asked to rate targets on 24 items mod-

eled after the 12 aspects of feedback provided on the SAPA-Project during the data collection period. This feedback was based on the 10 Big Five Aspect Scales (DeYoung et al., 2007) plus Honesty and Humility. Response choices for these informant report items ranged from 1 = “Very inaccurate” to 6 = “Very accurate”. Additionally, Informants were asked to judge the target’s physical attractiveness (Attractiveness), as well as math skills, verbal skills, and overall intelligence (these were indexed as “Rated IQ”). See [Table 1](#) for full item content. Unlike the self-report survey in the SAPA project, Informants were presented with all available informant-report items. After completing ratings of the target, informants were invited to complete the SAPA-Project survey for themselves, and if they chose to do so, their RID was used to identify their SAPA self-report. We identified 270 unique Informants who completed the SAPA survey in addition to providing ratings of the target.

During the time of this data collection, the self-report items administered to participants were drawn at random from a diverse pool of public-domain personality inventories; see Condon (2014, 2018) for more information. The personality feedback provided to participants during this period was based on the hypothesis that the evidence supporting use of the Big Five Aspect Scales (BFAS; DeYoung et al., 2007) could be extended to a six-factor model (Ashton et al., 2007; Thalmayer et al., 2011). The hypothesized model — with 12 aspects and 6 factors — was also used as a framework for informant-report data collection. This theory-based structure is visible in the item content shown in [Table 1](#). Subsequent to the data collection period, empirical evaluations of the structure of the self-report assessment model did not support the hypothesized framework. In fact, it was on the basis of these analyses that an alternative hierarchical model was first developed, the SAPA Personality Inventory (SPI, Condon, 2018).

The SPI is an empirically-derived, self-report personality assessment model with a hierarchical structure including both Big Five measures and a larger number (27) of more narrow, lower-order traits. As suggested above, the derivation of the SPI was based on the administration of nearly 700 personality items from dozens of (over-lapping) public-domain personality inventories to three large samples (combined  $N > 125,000$ ). All of these samples are included in the data used in the current study to report on self-informant correlations. The longest version of the SPI self-report measure — the one used in the current work — contains 135 items: each of the 27 lower-order traits is assessed with a 5 item scale, and each of the Big Five scales uses 14 items (only 70 of the 135 items are used for the Big Five scales). Unlike hierarchical personality models with fully “nested” traits at different levels (i.e., facets or aspects are nested under a specific Big Five domain), the 27 empirically-derived factors of the SPI are independent of the Big Five, meaning that several load highly on more than one higher order factor. For example, in the SPI model, the lower-level factor “Adaptability” is positively correlated with Extraversion (.38) and Openness (.39), and negatively correlated with Neuroticism (-.39). Full documentation of the SPI is given in Condon (2018), including a transparent walk-through of the code, sensitivity analyses, norms, and IRT parameters.

Here, we report mainly on self-informant associations of the SPI self-report model using the higher level SPI-5 scales (Agreeableness, Conscientiousness, Neuroticism, Extraversion, and Openness) along with the 27 lower-level scales. Comparisons with several other self-report frameworks are possible using the publicly available data at <https://dataverse.harvard.edu/dataverse/SAPA-Project>. To contextualize the validity of the SPI-5 relative to other frameworks, we report the self-informant associations for three of the more widely-used self-report frameworks: the BFAS (DeYoung et al., 2007), the IPIP correlates of the Big Five Factor Markers (IPIP5, Goldberg et al., 2006), and the IPIP correlates of the NEO-PI-R (IPIPNEO, Goldberg et al., 2006). The number of pairwise administrations between all of the self-report items in these frameworks (minimum = 1,988;  $Mdn = 2,770$ ;  $m = 3,128.1$ ) is considerably greater than the recommended number for stable correlational analyses (typically, 250 to 500; Kretzschmar & Gignac, 2019; Schönbrodt & Peurugini, 2013).

## Participants

All measures were administered in English. SAPA participants in the full self-report sample ( $N = 158,496$ ) represent 213 different countries, with 68% (93,271) of the sample residing in the United States. Over half the sample was female (61%). 1,554 informant-reports were received for 921 unique targets, of which 786 were matched to demographic information provided by participants who completed the self-report surveys. Targets resided in 64 countries, with roughly 67% (508) of Targets living in the United States. 65% of Targets were female.

Participants who completed an informant-report were assigned a 9-digit RID that did not persist beyond their web browser session; this was done to prevent the use of tracking software and potentially identifying log-ins. If an informant closed and reopened their web browser, they were assigned a new RID. Therefore, we can only estimate how many unique Informants completed the informant-reports. Out of the 1,554 reports received, only 30 Informants indicated they previously completed one of our informant-reports and thus, we estimate the total of unique Informants is approximately 1,500. The majority of Informants reported being friends of the target (55%), followed by romantic partners (12%), spouses (7%), siblings (7%), and parents (6%). 88% of Informants reported knowing the target “fairly well” or “extremely well.” Of the total number of Informants, 270 also completed the SAPA self-report. Additional demographic information is provided in [Table 2](#).

## Analysis

If a target was rated by the same informant more than once the most complete report was taken. If reports were similarly complete, the first report was taken. Of the 1,568 informant-reports, 14 reports were excluded for duplicated informant/target pairs, leaving 1,554 informant-reports for further analysis. Informant-reports were matched to self-reports with the last six-digits (Original Identification Number, OID) of the target’s RID. Twelve ratings were manually matched by time-stamp, as a small number of OIDs corresponded to more than one RID. Of the 921 unique tar-

**Table 1. Informant-report items and reliabilities**

Item text	Lower-level domain	Item-scale correlation
Agreeableness ( $\alpha = .47, \bar{r} = .18, Mdn r = .21$ )		
Is patient and polite	Polite	.51
Is indifferent to others' feelings	Compassion	-.72
Likes to help others	Compassion	.67
Tells people when they are frustrated	Polite	-.39
Conscientiousness ( $\alpha = .66, \bar{r} = .33, Mdn r = .31$ )		
Is disorganized	Orderly	-.78
Neglects their work/duties	Industry	-.77
Works hard	Industry	.68
Likes things to be "just right"	Industry	.59
Emotional Stability ( $\alpha = .75, \bar{r} = .43, Mdn r = .43$ )		
Is moody	Balanced	-.80
Is fearful	Boldness	-.80
Is composed	Balanced	.75
Faces danger confidently	Boldness	.65
Extraversion ( $\alpha = .62, \bar{r} = .29, Mdn r = .23$ )		
Is assertive, takes charge	Assertive	.71
Keeps others at a distance	Enthusiasm	-.69
Enjoys being with people	Enthusiasm	.68
Prefers to let others lead	Assertive	-.66
Intellect/Openness ( $\alpha = .53, \bar{r} = .22, Mdn r = .22$ )		
Can't handle a lot of information	Intellect	-.71
Is disinterested in abstract ideas	Openness	-.69
Understands things quickly	Intellect	.60
Believes in the importance of art	Openness	.59
Honesty/Humility ( $\alpha = .69, \bar{r} = .36, Mdn r = .33$ )		
Disregards rules to get ahead	Honesty	-.74
Shows off	Humility	-.73
Is humble	Humility	.72
Sticks to the rules	Honesty	.69
Rated IQ ( $\alpha = .50, \bar{r} = .25, Mdn r = .32$ )		
Has strong math skills	Cognitive ability	.77
Has strong verbal skills	Cognitive ability	.71
Is intelligent	Cognitive ability	.69
Attractiveness ( $\alpha = .89, \bar{r} = .74, Mdn r = .78$ )		
Is physically attractive	Attractiveness	.94
Has an attractive face	Attractiveness	.91
Has an attractive body.	Attractiveness	.88

Note. Item-scale correlations are corrected for item overlap and scale reliability. Negative correlations indicate reverse-coded items. The alpha ( $\alpha$ ), average  $r$  ( $\bar{r}$ ), and median  $r$  are provided for each scale.

gets, 786 were successfully matched to a RID in SAPA. All analyses with both self-reports and informant-reports included these 786 unique participants. Analyses of informant-reports alone maintained the sample of 921 targets. Item Response Theory (IRT) two parameter logistic scoring in the psych package *psych* (Revelle, 2018) in R (R Core Team, 2018) was used for both self-reports and informant-reports. These scores were used to compare the mean SPI-5

profiles of Openers, Senders, Targets, and Informants to the general respondent sample, and to identify the associations between the informant-report scales and the self-report scales, including the SPI measures as well as the BFAS, the IPIP5, and the IPIPNEO.

**Table 2. Demographics**

Size	Age	Gender	Percent US	Race/Ethnicity	Education
SAPA					
<i>n</i> = 158,496	<i>M</i> = 26 <i>SD</i> = 10.8 <i>Mdn</i> = 22	61% female	68% US	64% White 9% African American 9% Mexican American/Hispanic 6% Two or more 9% Other race/ethnicity 3% Not reported	14% less than 12yrs 9% high school degree 41% in college 6% some college (13-15yrs) 16% college degree 5% in graduate/prof school 9% graduate/prof degree
Openers					
<i>n</i> = 16,787	<i>M</i> = 27.9 <i>SD</i> = 12.1 <i>Mdn</i> = 24	64% female	63% US	69% White 6% African American 6% Mexican American/Hispanic 7% Two or more 9% Other race/ethnicity 3% Not reported	10% less than 12yrs 8% high school degree 37% in college 8% some college (13-15yrs) 19% college degree 6% in graduate/prof school 12% graduate/prof degree
Senders					
<i>n</i> = 4,380	<i>M</i> = 28.8 <i>SD</i> = 13 <i>Mdn</i> = 24	66% female	65% US	65% White 10% African American 7% Mexican American/Hispanic 7% Two or more 8% Other race/ethnicity 3% Not reported	12% less than 12yrs 8% high school degree 39% in college 7% some college (13-15yrs) 18% college degree 5% in graduate/prof school 11% graduate/prof degree
Targets					
<i>n</i> = 786	<i>M</i> = 28.7 <i>SD</i> = 12.3 <i>Mdn</i> = 24	65% female	67% US	71% White 4% African American 7% Mexican American/Hispanic 6% Two or more 9% Other race/ethnicity 3% Not reported	10% less than 12yrs 6% high school degree 40% in college 9% some college (13-15yrs) 19% college degree 5% in graduate/prof school 11% graduate/prof degree
Total Informants					
<i>n</i> = 1,500	<i>M</i> = 32.4 <i>SD</i> = 15.3 <i>Mdn</i> = 27	60% female			

Size	Age	Gender	Percent US	Race/Ethnicity	Education
Informants (with self-report data)					
<i>n</i> = 270	<i>M</i> = 35.8 <i>SD</i> = 16.4 <i>Mdn</i> = 30.5	68% female	67% US	78% White 5% African American 2% Mexican American /Hispanic 6% Two or more 6% Other race/ethnicity 3% Not reported	5% less than 12yrs 7% high school degree 19% in college 15% some college (13-15yrs) 32% college degree 3% in graduate/prof school 19% graduate/prof degree

*Note.* Race/ethnicity for US participants only.

Targets received ratings from between 1 and 21 informants; the modal number of reports received was 1. For targets with two or more reports, the mean value was taken for each item. Next, we constructed a correlation matrix of self-report and informant-report items. In order to mitigate the influence of replicated error, this matrix was scored adjusting for overlapping keys and replacing the overlapping covariances with the corresponding best estimate of the item's "true variance", the average correlation for that item following methods proposed by Bashaw & Anderson (1967) and Cureton (1966). If more than one informant rated a target, we used intraclass correlations to assess inter-informant agreement between the first two informant-reports for a given target.

The best items in SAPA for each informant-report scale were derived empirically using the `bestScales` function in the *psych* (Revelle, 2018) package in R (R Core Team, 2018). These procedures are described more fully by Elleman et al. (2020). Similar to the methods used for criterion-keyed scale construction, this method generates a unit-weighted model (for each criterion) based on the magnitudes of the zero-order correlations between the predictors (the items) and the informant-report scores following ten iterations of k-fold cross validation. The pool of self-report items used as predictors included all those in the SPI scales (135) and the remainder of the 696 items used during the empirical derivation of the SPI (Condon, 2018). A full listing of the items and all corresponding data are available in the public domain (Condon et al., 2017; Condon & Revelle, 2015). All of these analyses were exploratory (i.e., not preregistered).

## Results

### Group differences

Due to the large sample sizes collected, we discuss results in terms of effect size rather than statistical significance. Conventional criterion asserts Cohen's  $d$  of 0.2, 0.5, and 0.8, are small, medium, and large effect sizes respectively (Cohen, 1988). One meta-analytic review of individual differences research suggests small, medium, and large effects in empirical work are equivalent to Cohen's  $d$ s of 0.20, 0.41, 0.63 (Gignac & Szodorai, 2016). In line with both convention and empirical evidence we consider Cohen's  $d$  of 0.20 to be a small effect and Cohen's  $d$  of less than 0.20 to be a very small effect. Here we report only group differences of  $d = 0.10$  or greater (a very small effect), but advise readers to interpret results with attention to the size of the effect and 95% confidence intervals.

Analyses of the sub-group scores on the SPI-5 domains indicated that Openers ( $n = 16,787$ ) scored higher on openness than the typical SAPA participant,  $d = 0.12$ , 95% CI [0.11, 0.14]. Targets ( $n = 786$ ) were also more open than the general sample,  $d = 0.13$ , 95% CI [0.05, 0.20]. Targets who received two or more informant-reports ( $n = 282$ ) were more agreeable,  $d = 0.16$ , 95% CI [0.001, 0.32] and more open than targets with only one informant-report ( $n = 590$ ),  $d = 0.21$ , 95% CI [0.05, 0.37]. On average, Informants who completed the self-report survey ( $n = 270$ ) were more agreeable than the general SAPA sample,  $d = 0.18$ , 95% CI [0.04, 0.31].

Differences in Big Five traits between Senders and Targets (senders who received an informant-report) did not

meet our cut-off of Cohen's  $d = 0.10$  or greater (effect sizes ranged from 0.01 to 0.07). The absence of an effect in this case is noteworthy as it suggests that self-report personality ratings of the target do not affect whether informants agree to provide a rating. Taken together, the participants who opted-in to the informant-report system, either by sending a request or completing an informant-report, differed from typical participants in that they were more Open and Agreeable than the average SAPA respondent, although these were small effects. Respondents who received more than one report were even more Open and Agreeable compared to those with only one report, though the incremental effect was also small. See [Table 3](#) for descriptive statistics of the SPI-5 profiles for the general sample (SAPA), Openers, Senders, Targets and Informants.

### Self-informant agreement

Correlations between the informant- and self-report scales are shown in [Tables 4](#) and [5](#). In [Table 4](#), the correlations are corrected for reliability using the standardized alpha for each scale (Bashaw & Anderson, 1967; Cureton, 1966). Some informant scales showed low scale reliability, which can be likely attributed to their relatively short length (four items or fewer) and the intentional inclusion of content from two aspects for each domain (e.g., compassion and politeness in Agreeableness, DeYoung et al., 2007). However, the standardized alphas for three of the informant scales (indicated in [Table 1](#)) were particularly poor: Agreeableness (4 items,  $\alpha = .47$ ), Intellect/Openness (4 items,  $\alpha = .53$ ), and Rated IQ (3 items,  $\alpha = .50$ ). As low reliabilities can lead to large adjustments of the correlations when attenuated (Schmitt, 1996), we present the correlations without correcting for scale reliability in [Table 5](#).

In this sample, correlation values of .15 are significant at the .01 level using the Bonferroni correction for multiple comparisons. When corrected for scale reliability, the SPI-5 domains produced convergent validities between self- and informant-reports ranging from .42 to .72, with a median of .67. Among the 5 domains, discriminant validities ranged from .01 to .37 with median of .135; five values were above .20 and four of these were with Extraversion. Indeed, higher informant-ratings of Extraversion were positively associated with all Big Five self-ratings in the direction of social desirability. Raw convergent validities ranged from .28 to .55 ( $Mdn = .48$ ) and raw discriminant validities ranged from .01 to .27 ( $Mdn = .11$ ).

The structure of the self-informant agreement demonstrates the overlapping structure of the SPI measure ([Table 4](#)), where informant-ratings correlated with multiple lower-level SPI domains. For example, informant-reports of Agreeableness correspond with self-reported Compassion (.74), Trust (.51), Honesty (.36), and Irritability (-.39). Informant's judgments of Conscientiousness correspond with self-reported Order (.51), Industry (.65), Impulsivity (-.31), and Self-Control (.47). Likewise, the 27 SPI domains load on more than one informant-rating scale. For example, self-reported Well Being converges with the informant's rating of Conscientiousness (.28), Emotional Stability (.46), Extraversion (.56), and Intellect/Openness (.24). The raw correlations exhibited similar structural patterns ([Table 5](#)).

**Table 3. Descriptive statistics of the SPI-5 for all subsets**

Variable	<i>n</i>	<i>M</i>	<i>SE</i>	<i>SD</i>
<b>SAPA</b>				
Agreeableness	123819	0.09	0	0.99
Conscientiousness	133342	0.05	0	0.92
Extraversion	111884	0.06	0	0.44
Emotional Stability	110419	0.10	0	1.13
Openness	129392	-0.36	0	0.87
<b>Openers</b>				
Agreeableness	14422	0.07	0.01	0.93
Conscientiousness	15127	-0.02	0.01	0.89
Extraversion	13406	0.03	0.00	0.42
Emotional Stability	13360	0.09	0.01	1.09
Openness	14886	-0.26	0.01	0.79
<b>Senders</b>				
Agreeableness	3716	0.11	0.02	0.96
Conscientiousness	3878	0.04	0.01	0.90
Extraversion	3428	0.07	0.01	0.42
Emotional Stability	3423	0.12	0.02	1.13
Openness	3830	-0.29	0.01	0.82
<b>Targets</b>				
Agreeableness	706	0.12	0.03	0.90
Conscientiousness	709	-0.01	0.03	0.83
Extraversion	652	0.05	0.02	0.40
Emotional Stability	661	0.15	0.04	1.11
Openness	714	-0.25	0.03	0.78
<b>Informants</b>				
Agreeableness	215	0.26	0.06	0.84
Conscientiousness	229	0.03	0.06	0.92
Extraversion	180	0.07	0.03	0.45
Emotional Stability	193	0.08	0.08	1.12
Openness	214	-0.25	0.06	0.81

Note. Sample sizes vary within a given group for individual scales due to the “missingness” of the MMICAR data.

Targets and informants showed moderate agreement, especially on the Big-Five constructs; of these five, informants and targets agreed most strongly on the target’s level of Agreeableness (.72) and least strongly on the target’s Intellect and Openness (.42). Moreover, the structure of these agreements exhibit the same “heterarchical” (Milyavskaya et al., 2013) pattern proposed by the SPI model, where lower-level domains load on more than one higher-level domain in an overlapping sequence.

Table 6 shows self-informant agreement based on Big Five scores using the BFAS, IPIP5, and IPIPNEO, after correcting for scale reliability. With the exception of Openness, results are similar across the various operationalizations of the Big Five traits. Openness/Intellect associations were smaller than those for the other traits, and more disparate across models; the self-informant correlations were similar in magnitude for the BFAS and IPIP5 (.62 and .61, respectively), as were the lower correlations for the SPI and NEO

Openness (.42 and .49, respectively).

Two hundred and eighty-two targets received two or more informant-ratings and the agreement among raters is shown in Table 7. Informants agreed to a similar extent for all of the Big Five traits (*r*’s of .29 to .35) and Attractiveness ratings (.37), slightly less for Honest/Humility (.26), and not at all for judgments of intelligence (*ns*). Median inter-informant agreement was .34.

### Item-scale analysis

In order to understand what informants were responding to in their judgments of the target, we examined the self-report items that were the best predictors of informant-rating scores. Table 8 displays the self-report items that were consistently the best predictors of the informant-reports. The correlations of these “best items” with informant-report scales ranged from .23 to .56, with a median of .44.



**Table 4. Agreement between self and informant reports for the SPI-5 and 27 lower-level SPI domains**

Self-report	Informant report							
	Agree	Consc	EmoStab	Extra	IntelOpen	HonHum	RtdIQ	Attrac
SPI Agreeableness	<b>0.72</b>	0.20	0.15	<b>0.34</b>	0.18	<b>0.32</b>	0.14	<b>0.22</b>
SPI Conscientiousness	0.12	<b>0.68</b>	0.18	<b>0.36</b>	0.10	<b>0.24</b>	0.18	0.14
SPI Emotional Stability	-0.02	0.12	<b>0.67</b>	<b>0.37</b>	<b>0.22</b>	-0.14	0.19	-0.07
SPI Extraversion	0.02	-0.01	0.06	<b>0.63</b>	0.17	<b>-0.28</b>	0.07	0.12
SPI Openness	-0.10	-0.03	0.11	<b>0.20</b>	<b>0.42</b>	<b>-0.25</b>	<b>0.33</b>	-0.02
Compassion	<b>0.74</b>	0.19	0.01	<b>0.26</b>	<b>0.24</b>	<b>0.31</b>	0.19	<b>0.26</b>
Trust	<b>0.51</b>	0.05	0.18	<b>0.40</b>	0.12	0.09	0.07	0.18
Honesty	<b>0.36</b>	<b>0.27</b>	0.17	0.17	0.15	<b>0.36</b>	0.11	0.12
Conservatism	<b>0.47</b>	<b>0.35</b>	<b>0.26</b>	<b>0.24</b>	-0.07	<b>0.39</b>	0.15	0.09
Authoritarianism	<b>0.31</b>	<b>0.38</b>	0.08	0.19	-0.04	<b>0.43</b>	0.11	0.13
Easy Goingness	0.12	<b>-0.45</b>	-0.10	<b>-0.50</b>	<b>-0.24</b>	0.10	<b>-0.21</b>	-0.13
Perfectionism	-0.06	<b>0.35</b>	-0.12	0.18	0.10	0.01	<b>0.22</b>	0.11
Order	0.14	<b>0.51</b>	0.15	0.17	0.03	0.17	0.04	0.14
Industry	0.06	<b>0.65</b>	<b>0.25</b>	<b>0.44</b>	0.18	<b>0.23</b>	0.17	0.14
Impulsivity	-0.12	<b>-0.31</b>	<b>-0.29</b>	-0.13	-0.12	<b>-0.31</b>	-0.06	0.09
SelfControl	0.19	<b>0.47</b>	<b>0.38</b>	0.17	0.15	<b>0.34</b>	<b>0.24</b>	0.12
Emotional Stability	-0.16	0.18	<b>0.56</b>	<b>0.20</b>	0.16	-0.11	0.16	-0.15
Anxiety	0.19	-0.07	<b>-0.57</b>	<b>-0.40</b>	-0.17	<b>0.23</b>	-0.18	0.01
Irritability	<b>-0.39</b>	0.03	<b>-0.53</b>	-0.05	-0.08	-0.05	0.03	0.09
WellBeing	0.17	<b>0.28</b>	<b>0.47</b>	<b>0.58</b>	<b>0.27</b>	0.00	0.20	0.12
Emotional Expressiveness	0.18	0.10	-0.02	<b>0.37</b>	<b>0.22</b>	0.06	0.09	<b>0.26</b>
Sociability	0.19	0.01	0.05	<b>0.54</b>	0.06	-0.09	0.05	0.10
Adaptability	-0.11	-0.07	<b>0.24</b>	<b>0.32</b>	<b>0.26</b>	<b>-0.23</b>	0.09	0.00
Charisma	0.11	0.12	<b>0.21</b>	<b>0.67</b>	<b>0.30</b>	-0.19	0.12	0.19
Humor	0.18	-0.12	0.05	<b>0.41</b>	<b>0.26</b>	-0.10	0.15	0.14
Attention Seeking	<b>-0.27</b>	-0.10	0.00	<b>0.39</b>	0.13	<b>-0.48</b>	0.02	-0.02
Sensation Seeking	-0.15	-0.18	0.03	<b>0.26</b>	-0.03	<b>-0.49</b>	0.08	0.07
Conformity	<b>0.28</b>	0.15	0.03	0.07	<b>-0.26</b>	<b>0.35</b>	-0.10	0.07
Introspection	0.03	0.00	0.01	0.02	<b>0.29</b>	-0.10	<b>0.24</b>	-0.06
Art Appreciation	0.20	-0.03	0.10	0.05	<b>0.34</b>	0.04	0.07	0.19
Creativity	-0.13	-0.11	0.16	<b>0.26</b>	<b>0.34</b>	<b>-0.30</b>	<b>0.28</b>	0.00
Intellect	-0.15	<b>0.20</b>	0.11	<b>0.31</b>	<b>0.40</b>	-0.17	<b>0.32</b>	0.06

Note. Correlations are adjusted for reliability. See Table 5 for uncorrected correlations.

For each scale, the number of items ranged from 1 (Rated IQ) to 16 (Extraversion). These items provide insight into which self-reported aspects are being noticed and judged by the informant; for example, informant ratings of Agreeableness were best predicted by the targets' endorsement of items like "Cheer people up" and "Am sensitive to the needs of others." Likewise, targets' endorsement of items like "Am a person whose moods go up and down easily" and "Do things I later regret" were inversely related to the informant saying the target was emotionally stable. These results also corroborate a non-orthogonal structure for self-informant agreement; for example, "Find life difficult" is a strong predictor for informant reported Extraversion (inversely related, -.43) and Intellect/Openness (inversely re-

lated, -.28), suggesting this item taps into aspects of an informant's judgment of at least two domains. Notably, because these items are empirically-determined rather than theoretically-driven, they indicate which self-report items predict informant-reports irrespective of rationale; for example, "Am sensitive to the needs of others" is the self-report item that best predicts an informant's judgment of physical attractiveness even though these concepts are not clearly linked. In general, these results exhibit the potential of item-scale analysis to bridge the gap between targets' phenomenological self-perception and informants' perceptions/judgments of manifest personality.

**Table 5. Raw correlations between self-reports and informant-reports.**

Self-report	Informant report							
	Agree	Consc	EmoStab	Extra	IntelOpen	HonHum	RtdIQ	Attrac
SPI Agreeableness	<b>0.48</b>	0.15	0.12	<b>0.25</b>	0.13	<b>0.25</b>	0.10	0.20
SPI Conscientiousness	0.08	<b>0.52</b>	0.14	<b>0.27</b>	0.07	0.19	0.12	0.12
SPI Emotional Stability	-0.02	0.10	<b>0.55</b>	<b>0.27</b>	0.15	-0.11	0.14	-0.06
SPI Extraversion	0.01	-0.01	0.05	<b>0.48</b>	0.12	<b>-0.22</b>	0.05	0.11
SPI Openness	-0.07	-0.02	0.08	0.15	<b>0.28</b>	-0.19	<b>0.23</b>	-0.02
Compassion	<b>0.50</b>	0.15	0.01	0.19	0.16	<b>0.24</b>	0.14	<b>0.23</b>
Trust	<b>0.34</b>	0.04	0.14	<b>0.30</b>	0.08	0.07	0.05	0.16
Honesty	<b>0.23</b>	<b>0.20</b>	0.14	0.12	0.10	<b>0.27</b>	0.07	0.10
Conservatism	<b>0.30</b>	<b>0.25</b>	0.19	0.17	-0.05	<b>0.29</b>	0.10	0.08
Authoritarianism	<b>0.20</b>	<b>0.29</b>	0.06	0.13	-0.03	<b>0.32</b>	0.08	0.11
Easy Goingness	0.07	<b>-0.31</b>	-0.07	<b>-0.33</b>	-0.15	0.07	-0.13	-0.10
Perfectionism	-0.04	<b>0.24</b>	-0.08	0.12	0.06	0.01	0.14	0.08
Order	0.09	<b>0.38</b>	0.12	0.12	0.02	0.13	0.03	0.12
Industry	0.04	<b>0.49</b>	<b>0.20</b>	<b>0.32</b>	0.12	0.17	0.12	0.12
Impulsivity	-0.08	<b>-0.24</b>	<b>-0.23</b>	-0.09	-0.08	<b>-0.24</b>	-0.04	0.08
Self Control	0.12	<b>0.34</b>	<b>0.28</b>	0.12	0.10	<b>0.25</b>	0.16	0.10
Emotional Stability	-0.10	0.13	<b>0.44</b>	0.15	0.11	-0.09	0.11	-0.13
Anxiety	0.13	-0.06	<b>-0.46</b>	<b>-0.29</b>	-0.12	0.18	-0.13	0.01
Irritability	<b>-0.26</b>	0.02	<b>-0.43</b>	-0.04	-0.05	-0.04	0.02	0.08
Well Being	0.12	<b>0.22</b>	<b>0.38</b>	<b>0.44</b>	0.19	0.00	0.14	0.11
Emotional Expressiveness	0.12	0.08	-0.01	<b>0.26</b>	0.14	0.04	0.06	<b>0.22</b>
Sociability	0.12	0.01	0.04	<b>0.39</b>	0.04	-0.07	0.03	0.09
Adaptability	-0.07	-0.05	0.18	<b>0.22</b>	0.17	-0.17	0.06	0.00
Charisma	0.07	0.09	0.16	<b>0.48</b>	0.20	-0.14	0.09	0.16
Humor	0.12	-0.09	0.04	<b>0.29</b>	0.17	-0.07	0.10	0.12
Attention Seeking	-0.18	-0.08	0.00	<b>0.29</b>	0.09	<b>-0.38</b>	0.02	-0.02
Sensation Seeking	-0.10	-0.14	0.03	0.19	-0.02	<b>-0.38</b>	0.05	0.06
Conformity	0.18	0.11	0.03	0.05	-0.17	<b>0.26</b>	-0.07	0.06
Introspection	0.02	0.00	0.01	0.02	0.19	-0.08	0.16	-0.05
Art Appreciation	0.13	-0.02	0.08	0.03	<b>0.22</b>	0.03	0.05	0.16
Creativity	-0.08	-0.08	0.13	0.19	<b>0.23</b>	<b>-0.23</b>	0.20	0.00
Intellect	-0.10	0.15	0.09	<b>0.22</b>	<b>0.27</b>	-0.13	<b>0.23</b>	0.05

Note. Correlations are not adjusted for reliability. See Table 4 for corrected correlations.

## Discussion

Self-reports of anonymous web users showed strong correspondence with the ratings of anonymous friends, family, colleagues, and other informants. This correspondence was seen at three levels of analysis: the Big Five domains, the 27 lower-level SPI factors, and at the item-scale level. We demonstrate the efficacy, as well as feasibility, of collecting large samples of informant-reports for free online. The more detailed structure of the self reports provided by the 27 sub-scales of the SPI provides validity for both the broader and narrower domains. In addition, by using SAPA procedures to collect data from 696 items, we were also able to examine the pattern of individual item correlations with

the informant ratings, which provides more detailed information about the aspects of the individual's identity being rated by informants.

Self-informant agreement on the SPI-5 domains ranged from .42 (Intellect/Openness) to .72 (Agreeableness), with a median of .67, when adjusted for scale reliability and item overlap. Our results are similar to those found for the levels of self-peer consensus on both 60 and 30-item measures of the Big Five (Konstabel et al., 2017). Contrary to prior work (Norman, 1969; Vazire, 2010), we found no evidence for substantially lower self-informant agreement on Emotional Stability/Neuroticism (.67) than in other domains. In this study, informants were selected by the target, and consequently, the majority of informants, regardless of the re-

**Table 6. Self-informant agreement with the Big Five Aspect Scales (BFAS), the IPIP correlates of the Big Five Factor Markers (IPIP5), and the NEO-PI-R (IPIPNEO)**

Self-report	Informant report							
	Agree	Consc	EmoStab	Extra	IntelOpen	HonHum	RtdIQ	Attrac
BFAS Agreeableness	<b>0.78</b>	0.16	0.06	0.17	0.14	<b>0.50</b>	0.06	<b>0.22</b>
BFAS Conscientiousness	0.09	<b>0.66</b>	<b>0.28</b>	<b>0.30</b>	0.16	<b>0.28</b>	0.11	0.10
BFAS Neuroticism	<b>-0.21</b>	-0.20	<b>-0.66</b>	<b>-0.36</b>	-0.18	-0.05	-0.06	0.08
BFAS Extraversion	0.13	0.16	0.18	<b>0.70</b>	<b>0.24</b>	-0.16	0.14	<b>0.28</b>
BFAS Openness	0.05	0.14	<b>0.21</b>	<b>0.25</b>	<b>0.62</b>	-0.10	<b>0.37</b>	0.13
IPIP5 Agreeableness	<b>0.65</b>	0.11	0.10	<b>0.37</b>	0.18	<b>0.30</b>	0.06	<b>0.25</b>
IPIP5 Conscientiousness	0.12	<b>0.63</b>	<b>0.21</b>	0.19	0.03	<b>0.28</b>	0.10	0.07
IPIP5 Emotional Stability	0.17	0.16	<b>0.70</b>	<b>0.35</b>	0.16	-0.02	0.03	-0.08
IPIP5 Extraversion	0.09	0.12	<b>0.21</b>	<b>0.74</b>	<b>0.23</b>	-0.15	0.16	0.18
IPIP5 Intellect	-0.03	0.06	<b>0.22</b>	<b>0.31</b>	<b>0.61</b>	-0.17	<b>0.39</b>	0.04
IPIPNEO Agreeableness	<b>0.77</b>	0.12	0.15	0.16	0.11	<b>0.46</b>	0.05	0.18
IPIPNEO Conscientiousness	0.17	<b>0.61</b>	<b>0.35</b>	<b>0.39</b>	<b>0.27</b>	<b>0.27</b>	<b>0.28</b>	0.14
IPIPNEO Neuroticism	-0.09	<b>-0.29</b>	<b>-0.63</b>	<b>-0.50</b>	<b>-0.29</b>	0.05	<b>-0.21</b>	-0.03
IPIPNEO Extraversion	0.11	0.11	0.19	<b>0.76</b>	<b>0.22</b>	<b>-0.20</b>	0.18	<b>0.20</b>
IPIPNEO Openness	0.01	-0.17	0.04	0.14	<b>0.49</b>	<b>-0.23</b>	0.16	0.07

Note. Correlations are adjusted for reliability.

**Table 7. Inter-informant agreement**

	Agree	Consc	EmoStab	Extra	IntelOpen	HonHum	RtdIQ	Attrac
Agreeableness	<b>0.34</b>							
Conscientiousness	0.08	<b>0.35</b>						
Emotional Stability	0.09	0.06	<b>0.34</b>					
Extraversion	0.15	0.15	0.15	<b>0.35</b>				
Intellect/Openness	0.06	0.04	0.08	0.04	<b>0.29</b>			
Honesty Humility	0.19	0.10	0.02	-0.03	0.01	<b>0.26</b>		
RatedIQ	-0.02	0.00	0.00	0.04	0.12	-0.01	0.06	
Attractiveness	0.10	0.07	0.00	0.05	-0.01	0.09	0.01	<b>0.37</b>

Note. For targets rated more than once, the first two informant ratings were compared using intraclass correlations using the ICC2k function from the *psych* (Revelle, 2018) package in R (R Core Team, 2018)

ported relationship with the target, reported knowing the target fairly or extremely well. This high level of acquaintanceship could be a possible explanation for the higher than expected self-informant agreement on Emotional Stability. Consistent with this explanation, Vazire reports higher levels of agreement between self and friends on Neuroticism traits (Self-esteem and Anxiety) compared to the agreement between the self and a stranger (Vazire, 2010). However, there is mixed evidence that self-other agreement across various traits, including Emotional Stability (Neuroticism), increases with greater acquaintanceship (Biesanz et al., 2007; Norman, 1969).

More generally, our results are consistent with predictions suggested by Vazire’s Self-Other Knowledge Asymmetry model (Vazire, 2010) whereby the self and other are more strongly aligned on highly visible traits (i.e., Agreeableness,  $r = .72$  in our sample) than less visible traits (i.e., Intellect and Openness,  $r = .42$  in our sample). Self-informant agreement on Extraversion ( $r = .58$ ) was somewhat lower than anticipated relative to the other domains, as Extraversion has shown the highest self-other agreement among the Big Five in prior work (Konstabel et al., 2017; Norman, 1969; Vazire, 2010). Given that reliability of the four-item informant-report Extraversion scale is adequate

**Table 8. Self-report items that best predict informant-reports**

Item	Item-scale correlation
Agreeableness (.46)	
Cheer people up.	.56
Am sensitive to the needs of others.	.48
Can't be bother with others' needs.	-.47
Sympathize with others' feelings.	.46
Am concerned about others.	.45
Tend to dislike soft-hearted people.	-.45
Am not interested in other people's problems.	-.45
Like to do things for others.	.44
Am inclined to forgive others	.44
Don't understand people who get emotional.	-.44
Will do anything for others.	.44
Criticize others' shortcomings.	-.42
Inquire about others' well-being	.42
Am indifferent to feelings of others	-.42
Conscientiousness (.44)	
Complete tasks successfully.	.47
Work hard	.45
Find it difficult to get down to work.	-.45
Feel desperate.	-.40
Make careless mistakes.	-.39
Do just enough work to get by.	-.38
Emotional Stability (.45)	
Am a person whose moods go up and down easily.	-.49
Do things I later regret.	-.48
Get upset by unpleasant thoughts that come into my mind.	-.45
Find life difficult.	-.44
Am not easily annoyed.	.44
Panic easily.	-.44
Am relaxed most of the time.	.44
Get caught up in my problems	-.42
Get upset easily.	-.41
Extraversion (.52)	
Tend to keep in the background on social occasions.	-.56
Avoid company.	-.50
Cheer people up.	.49
Feel at ease with people.	.49
Other people think of me as being very lively.	.49
Prefer to be alone.	-.48
Am skilled in handling social situations	.48
Have many friends.	.47
Keep in the background.	-.47
Usually enjoy being with people.	.46
Keep others at a distance.	-.45
Talk to a lot of different people at parties.	.44
Don't mind being the center of attention.	.44
Feel comfortable around people.	.44

Only feel comfortable with friends.	-43
Find life difficult.	-43
Intellect/Openness (.35)	
Don't understand things.	-43
Am quick to understand things.	.34
Can't make up my mind.	-34
Worry too long after an embarrassing experience.	-33
Recover quickly from stress and illness.	.29
Find life difficult.	-28
Honesty/Humility (.36)	
Make careless mistakes.	-43
It is better to follow society's rules than to go my own way.	.39
Make myself the center of attention.	-37
Enjoy practical jokes that can sometimes really hurt people.	-36
Would never make a high-risk investment.	.35
Do crazy things.	-33
Like to attract attention.	-33
Rated IQ (.23)	
Turn plans into actions.	-23
Attractiveness (.33)	
Am sensitive to the needs of others.	.40
There are several people who keep trying to avoid me.	-33
Criticize others' shortcomings.	-33
Suffer from others' sorrows.	.32

Note. Mean item-scale correlations displayed in parentheses.

(i.e., not problematic), we speculate that the lower self-informant agreement may be attributed to a mismatch in item content between Extraversion self-report and informant-report scales. Specifically, informants rated targets on Extraversion with respect to the target's assertiveness and social enthusiasm, completing items such as "Prefers to let others lead" (assertiveness, negative) and "Enjoys being with people" (enthusiasm, positive.) In contrast, the self-report items in SPI Extraversion included attention-seeking items such as "Like to attract attention" (attention-seeking, positive), and did not include assertiveness items. This explanation is supported further by the best items analyses as the empirically-derived scale of self-report items is more highly associated with informant ratings for Extraversion (.52) than any other Big Five domain (.35 to .46).

It is also the case that informant-ratings of Extraversion were significantly associated with self-report ratings for all of the Big Five. This was unique to Extraversion — none of the other informant-report Big Five ratings had significant associations with self-reports of other traits except Intellect/Openness (only with Emotional Stability). These associations were also substantial and uniformly in the socially desirable direction. In other words, higher informant ratings of Extraversion was correlated with higher *self-ratings* of Agreeableness, Conscientiousness, Emotional Stability, and Intellect/Openness (and Extraversion). We interpret this to suggest that informants perceive — and perhaps conflate — socially desirable behavioral tendencies as being

more extraverted.

There are at least two important aspects of the findings with respect to the consistency of ratings across informants. The first relates to the magnitude of the rates of agreement. Agreeableness, Extraversion, Emotional Stability, and Conscientiousness were nearly identical (.34-.35), with Intellect/Openness and Honesty/Humility slightly lower (.29 and .26, respectively). These magnitudes can be contextualized relative to the agreement for physical attractiveness, which is at the high end of this same range (.37). This contextualization is meaningful because physical attractiveness is often noted to be highly consensual (Ibáñez-Berganza et al., 2019). Though studies focused more directly on evaluations of attractiveness tend to report higher effect sizes (i.e., meta-analyses show mean inter-rater reliabilities of .21-.71; Langlois et al., 2000), these typically account for multiple moderators including age and cultural factors. The evidence for nearly identical consistency in Big Five informant-ratings suggests roughly equivalent consensus for these constructs. The second interesting finding from these analyses was the *absence* of statistically significant consistency in informant-ratings of intelligence (.06, *ns*). Inter-informant ratings between perceived intelligence and intellect/openness were higher in magnitude (.12), though also not significant. It is unlikely that these findings were caused by low internal consistency among the perceived intelligence items as informant scales with similar alphas (Agreeableness, Intellect/Openness) were not similarly affected.

In contrast to other recent work on perceptions of intelligence (A. J. Lee et al., 2017; Stellar & Willer, 2018), we did not find consensus among informants in this sample with respect to targets' math skills, verbal skills, and overall intelligence, and this was unique among the traits assessed.

When compared to the general online respondent sample, participants who interacted with the opt-in system were generally more agreeable and open. Participants who opened the request page and those who were the target of informant-ratings scored higher on openness than the general sample. Additionally, Informants (who later completed a self-report measure) were more agreeable than the average respondent. It is unclear, however, if this could be generalized to the entire group of Informants (those with and without self-reports). The informant-report system itself has elements of both self- and other-selection, where an individual must both opt-in (self-selection) and receive a rating (other-selection) to become a target. Due to the self-selected and other-selected nature of this informant-rating system, it is likely the traits of both the prospective target and the prospective informant(s) played *some* role in the exchange of requests and completed informant-ratings, even though there were no substantial differences in Big Five traits for those targets who were rated by informants and those who only solicited ratings (but were not rated). In other words, the fact that our available sample of self-informant reports is *not representative* of the larger SAPA sample is noteworthy and could belie how individual traits interact with the collection of self and informant reports where the system is both optional and sans compensation.

When comparing self-informant agreement of the SPI to the other operationalizations of the Big Five (BFAS, IPIP5, and IPIPNEO), the findings are more similar than different. The most prominent differences are for Openness/Intelligence, where the SPI and IPIPNEO self-reports are both less highly associated with the informant reports than the BFAS and IPIP5. These pairings are unsurprising given the derivation of the BFAS from the IPIP Big Five (DeYoung et al., 2007), but the lower magnitudes — especially for SPI and NEO Openness — reflect the reality that longer self-report measures of this trait in particular include more diverse content than is covered by the 4-item informant-report. The SPI Extraversion self-informant correlation is also somewhat lower than the other models (.63 vs .70-.76), but this can once again be explained by the lack of empirical evidence supporting the inclusion of Assertiveness in the derivation of the SPI (Condon, 2018). In fact, this is an important theme of the findings reported here as there was considerable content included in the pools of items used to derive the SPI, including all of the content from the BFAS, IPIP5, and the IPIPNEO (Condon, 2018).

### Limitations and future directions

In this study, we discuss the convergence of self and informant report items, but it should be reiterated the self- and informant-report assessment models were dissimilar by design. As discussed above, the informant survey was based on a theory-driven “aspects” model of the Big Five/Big Six domains whereas the self-report instrument was derived empirically based on the structure underlying a large pool

of personality items. This is a limitation because it led to an imperfect evaluation of self- and informant-ratings based on common structure. This means that the results were conflated beyond the simple self/other distinction because we are also mapping across different frameworks. Future research should consider a more direct evaluation of the relations between self- and informant-ratings using an informant-report model that matches the self-report SPI model more directly. Note that using 2-item observer report scales for 27 factors would more than double the burden on observers relative to the 24 items administered here.

Despite the benefits of an online, opt-in informant-report system, the rate of data collection is slow compared to other collection methods. We collected 1,554 informant-reports over 1,335 days, suggesting an average return rate of approximately one informant-report per day. Compare this to the average collection rate of 117 SAPA self-reports per day over the same time period. Although this study demonstrates the viability of a large, global sample of informant-reports, the length of time needed to obtain this large sample may be considered untenable for some researchers. Future research could explore ways to improve the rate of informant-report completion. For instance, the SAPA self-report survey is founded on the premise that participants are motivated to complete the assessment to “know thyself,” and participants are provided feedback on their personality rather than compensated outright. In line with this premise, perhaps informants could be similarly motivated to know more about their relationship to the target. On the basis of the results reported here, we think the pursuit of initiatives like these is well-justified, and it seems especially promising to consider the use of various “gamification” strategies (Keusch & Zhang, 2017).

For example, researchers could increase informants' motivation to complete the report by supplying the informant and target feedback on how closely their responses are aligned. In conjunction with participant feedback, the rate of data collection could be improved by increasing the overall proportion of participants who request informant-reports and expand the pool of potential informants for one target. In the current study, only 3% of the sample chose to send a informant-report request. Requests were sent via email to specified recipients (i.e., the sender sends the link directly to a potential informant). This method requires the participant to know the potential informant well-enough to have their email address, which may be why the vast majority of Informants (88%) reported knowing the target fairly or extremely well. This method also requires the participant to invite each potential informant individually meaning participant burden increases linearly with the number of potential informants. Alternatively, the informant-report system could be linked to popular social media sites, whereby a participant could broadcast a request to a large number of unspecified recipients. This approach would lessen the burden of sending multiple requests and could expand the pool of potential informants for a single target. This one-to-many approach could also improve the viability of examining convergence by how closely the informant knows the target, an avenue that goes unexplored in the current study due to low variability in target-informant intimacy.

The convergence of informant-reports and self-reports demonstrates the efficacy of collecting informant data broadly and online. We demonstrate that this agreement can be accessed at various levels of analysis, from large domains to the item-level. Additionally, sufficiently-powered item-scale analyses are particularly informative for identifying which aspects of the target's self-reported identity are being rated by informants. Although we recognize the design and scope of the SAPA Project is unique, the informant-report methodology used here is more traditional in design (in that informant reports do not use planned missingness) and more accessible for other researchers (because the targets are responsible for soliciting informants). Thus, we hope that this aspect of the SAPA Project can provide insight into the power of collecting informant-reports and the feasibility of doing so at scale via the web. As assessment moves increasingly online, researchers have the opportunity to overcome a reliance on self-reports alone and utilize the internet to implement large, multi-method designs.

---

### Contributions

Contributed to conception and design: DMC, WR

Contributed to acquisition of data: DMC, WR  
Contributed to analysis and interpretation of data: AZ, DMC, WR  
Drafted and/or revised the article: AZ, DMC, WR  
Approved the submitted version for publication: AZ, DMC, WR

### Competing interests

The authors declare they have no significant competing financial, professional, or personal interests that might have influenced the performance or presentation of the work described in this manuscript.

### Data accessibility

The data are publicly available online at <https://dataverse.harvard.edu/dataverse/SAPA-Project>. The R script, which includes code for downloading the data directly into the R environment through an API, is available online as supplementary material.

Accepted: July 16, 2021 PDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

## REFERENCES

- 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. <https://doi.org/10.1016/j.paid.2006.10.027>
- Bashaw, W. L., & Anderson, H. E. (1967). A correction for replicated error in correlation coefficients. *Psychometrika, 32*(4), 435–441. <https://doi.org/10.1007/BF02289657>
- Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology, 92*(1), 119. <https://doi.org/10.1037/0022-3514.92.1.119>
- Clifton, A., Turkheimer, E., & Oltmanns, T. F. (2005). Self-and peer perspectives on pathological personality traits and interpersonal problems. *Psychological Assessment, 17*(2), 123. <https://doi.org/10.1037/1040-3590.17.2.123>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.
- Condon, D. M. (2014). *An organizational framework for the psychological individual differences: Integrating the affective, cognitive, and conative domains*. Northwestern University.
- Condon, D. M. (2018). The SAPA Personality Inventory: An empirically-derived, hierarchically-organized self-report personality assessment model. *PsyArXiv*, 1–444. <https://doi.org/10.31234/osf.io/sc4p9>
- Condon, D. M., & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence, 43*, 52–64. <https://doi.org/10.1016/j.intell.2014.01.004>
- Condon, D. M., & Revelle, W. (2015). Selected personality data from the SAPA-Project: On the structure of phrased self-report items. *Journal of Open Psychology Data, 3*(8), 1515–1516. <https://doi.org/10.5334/jopd.al334>
- Condon, D. M., Roney, E., & Revelle, W. (2017). A SAPA Project update: On the structure of phrased self-report personality items. *Journal of Open Psychology Data, 5*(1). <https://doi.org/10.5334/jopd.32>
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*(6), 1092. <https://doi.org/10.1037/a0021212>
- Cureton, E. E. (1966). Corrected item-test correlations. *Psychometrika, 31*(1), 93–96. <https://doi.org/10.1007/BF02289461>
- 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. <https://doi.org/10.1037/0022-3514.93.5.880>
- Elleman, L. G., McDougald, S. K., Condon, D. M., & Revelle, W. (2020). That takes the BISCUIT: Predictive accuracy and parsimony of four statistical learning techniques in personality data, with data missingness conditions. *European Journal of Psychological Assessment, 36*(6), 948–958. <https://doi.org/10.1027/1015-5759/a000590>
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>
- 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, pp. 7–28). 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. <https://doi.org/10.1016/j.jrp.2005.08.007>
- Hogan, R. (2007). *Personality and the fate of organizations*. Lawrence Erlbaum Associates Publishers.
- Ibáñez-Berganza, M., Amico, A., & Loreto, V. (2019). Subjectivity and complexity of facial attractiveness. *Scientific Reports, 9*(1), 1–12. <https://doi.org/10.1038/s41598-019-44655-9>
- Internet World Stats. (2021). *Internet Usage Statistics: The Internet Big Picture*. <https://www.internetworldstats.com/stats.htm>
- Keusch, F., & Zhang, C. (2017). A review of issues in gamified surveys. *Social Science Computer Review, 35*(2), 147–166. <https://doi.org/10.1177/0894439315608451>
- Konstabel, K., Lönnqvist, J.-E., Leikas, S., Velazquez, R. G., Qin, H., Verkasalo, M., & Walkowitz, G. (2017). Measuring single constructs by single items: Constructing an even shorter version of the “Short Five” personality inventory. *PloS One, 12*(8), e0182714. <https://doi.org/10.1371/journal.pone.0182714>
- Kretzschmar, A., & Gignac, G. E. (2019). At what sample size do latent variable correlations stabilize? *Journal of Research in Personality, 80*, 17–22. <https://doi.org/10.1016/j.jrp.2019.03.007>
- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*(3), 390. <https://doi.org/10.1037/0033-2909.126.3.390>



- Lee, A. J., Hibbs, C., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2017). Assessing the accuracy of perceptions of intelligence based on heritable facial features. *Intelligence*, *64*, 1–8. <https://doi.org/10.1016/j.intell.2017.06.002>
- Lee, K., & Ashton, M. C. (2017). Acquaintanceship and self/observer agreement in personality judgment. *Journal of Research in Personality*, *70*, 1–5. <https://doi.org/10.1016/j.jrp.2017.05.001>
- McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The trait-reputation-identity model. *Psychological Review*, *123*(5), 569–591. <https://doi.org/10.1037/rev0000035>
- Milyavskaya, M., Philippe, F., & Koestner, R. (2013). Psychological need satisfaction across levels of experience: Their organization and contribution to general well-being. *Journal of Research in Personality*, *47*, 41–51. <https://doi.org/10.1016/j.jrp.2012.10.013>
- Norman, W. T. (1969). “To see ourselves as others see us!”: Relations among self-perceptions, peer-perceptions, and expected peer-perceptions of personality attributes. *Multivariate Behavioral Research*, *4*(4), 417–443. [https://doi.org/10.1207/s15327906mbr0404\\_1](https://doi.org/10.1207/s15327906mbr0404_1)
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston. <https://CRAN.R-project.org/package=psych>
- Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A., & Elleman, L. G. (2016). Web and phone based data collection using planned missing designs. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *SAGE Handbook of Online Research Methods* (2nd ed., pp. 578–595). Sage Publications, Inc.
- Schmitt, N. (1996). Uses and Abuses of Coefficient Alpha. *Psychological Assessment*, *8*(4), 350–353. <https://doi.org/10.1037/1040-3590.8.4.350>
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. <https://doi.org/10.1016/j.jrp.2013.05.009>
- Stellar, J. E., & Willer, R. (2018). Unethical and inept? The influence of moral information on perceptions of competence. *Journal of Personality and Social Psychology*, *114*(2), 195. <https://doi.org/10.1037/pspa000097>
- 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. <https://doi.org/10.1037/a0024165>
- Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, *40*(5), 472–481. <https://doi.org/10.1016/j.jrp.2005.03.003>
- Vazire, S. (2010). Who knows what about a person? The self--other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281–300. <https://doi.org/10.1037/a0017908>

## SUPPLEMENTARY MATERIALS

### Peer Review History

Download: [https://collabra.scholasticahq.com/article/25983-the-convergence-of-self-and-informant-reports-in-a-large-online-sample/attachment/65867.docx?auth\\_token=zi8NwMByda8umkHbKwud](https://collabra.scholasticahq.com/article/25983-the-convergence-of-self-and-informant-reports-in-a-large-online-sample/attachment/65867.docx?auth_token=zi8NwMByda8umkHbKwud)

---

### R script

Download: [https://collabra.scholasticahq.com/article/25983-the-convergence-of-self-and-informant-reports-in-a-large-online-sample/attachment/65868.zip?auth\\_token=zi8NwMByda8umkHbKwud](https://collabra.scholasticahq.com/article/25983-the-convergence-of-self-and-informant-reports-in-a-large-online-sample/attachment/65868.zip?auth_token=zi8NwMByda8umkHbKwud)

---