Pre-registration Protocol: Smartphone Sensing Panel Study - Personality Traits, Life Outcomes and Behavior

This pre-registration protocol deals with specific research questions and is completed before the data is accessed. Throughout this registration, we will refer to the corresponding basic registration protocol of the panel study. The basic protocol contains information on study procedures and further background information and can be found in the general pre-registration template here: https://doi.org/10.23668/PSYCHARCHIVES.2901. This template was inspired by the OSF Prereg Challenge template (https://osf.io/).

Working Title
The Predictiveness of Personality Traits and Behavioral Sensing Data for Life Outcomes

Author(s) of the preregistration protocol
Clemens Stachl, Stanford University

Date
05.06.2020

Background

Background Information (Optional; Short description of the theoretical background/introduction to research question)
Short summary: In this project we will investigate and compare the predictiveness of personality traits and objective sensing data from smartphones for life outcomes. Predicting life outcomes from Mobile Sensing Data Personality traits are generally described as relatively stable patterns of thought, feelings, and behaviors that are relevant in many parts of life. Further, personality traits are important because

they predict important life outcomes (Ozer & Benet-Martínez, 2006, Roberts, Kuncel, Shiner, Caspi & Goldberg, 2007 Soto, 2019). Hence, knowing someone's personality is extremely useful to assess whether a person would be open to discuss ideas over a coffee, reliable to work with, good to have as a friend, or good to party with. While these findings are impressive, they are somewhat limited because they entirely rely on self-reported data from questionnaires and experience sampling and on in-sample correlations or regression modeling. The fundamental problems of self-reports have been known for a long time, especially with regard to behavioral data (Ellis, Davidson, Shaw & Geyer, 2019, Furr, 2009, Gosling, John, Craik & Robins, 1998, Paulhus & Vazire, 2007). Until recently, it was extremely difficult to collect large amounts of objective data on how people actually behave.

However, with the advent of new computing technologies in the last two decades, researchers now have the tools to collect and analyze data from objective quantifications of individual differences to determine the role psychological traits and states play for life outcomes (Boyd, Pasca & Lanning, 2020). Especially promising are behavioral and situational data that can be gathered from smartphones (Harari, 2015, 2016, Miller, 2012). Further, it has been shown that these data are both associated with and predictive of personality traits at the factor and facet level (Harari, et al., 2019, Stachl et al., 2017, 2019). The richness, unobtrusiveness, objectivity and fine granularity of these data increasingly raises the question of whether self-reports should still be considered the gold standard or ground truth for the quantification of personality traits and individual differences (Boyd et al., 2020, Boyd & Pennebaker, 2017).
In particular this question is relevant if prediction rather than explanation is the primary goal.

The predictiveness of personality traits (for behavior and life outcomes alike) has been frequently praised in the literature (Ozer & Benet-Martínez, 2006, Roberts, Kuncel, Shiner, Caspi & Goldberg, 2007 Soto, 2019). However, predictive modeling (i.e., out-of-sample model evaluation) has rarely been used in past research to evaluate this claim. While, similar in-sample associations might be observable across replicating studies, it will be more interesting to see if models for the prediction of life outcomes can be created based on self-reports of personality and in-vivo behaviors and how well the models generalize beyond individual samples (Soto, in press).

This motivates us to compare how well state of the art self-report measures of psychological traits (Big Five personality traits), and sensing data from smartphones can be associated with, and predict life outcomes.

Research question(s)

Primary

The predictiveness of personality self-reports and objective behavioral sensing data for life outcomes

The main research question of this study is to determine how self-reported personality trait measures (self-reported personality traits) and objective behavioral sensing data (mobile sensing) compare in terms of how well they can predict life outcome measures (out-of-sample, cross-validated evaluation). See the “Variables” section and the basic pre-registration protocol for details on these measures. This analysis will evaluate whether personality traits at different levels

can predict life outcomes and whether behavioral sensing data can also be used to achieve this goal (H1a-b).

Additionally, we will evaluate the combined predictive effect of the combination of self-reported personality traits and behavioral sensing data from smartphones. While psychometric measures (self-reports) are purposely designed to be unidimensional (i.e. each item is uniquely related to only one psychological construct), the features derived from sensing data we will use in this study are not. Hence, one and the same type of sensing feature from smartphones can be indicative of various behaviors, situations, affect, and personality (state and trait). As manually labeling these features would be highly subjective and also infeasible, we think that the most practical way forward is to use computational methods (machine learning) for modeling the data.
At the same time, items in personality questionnaires are designed to cover all aspects of personality traits (thoughts, feelings, behaviors). Hence, it is to be determined whether both types of predictors (personality self-reports, behavioral sensing data) score better by itself (H2) or if the combination of both types of predictors will improve predictive performance for life outcomes overall (e.g., Settanni et al., 2018; H5).

Secondary
Replication of other previously reported effects
How strongly are life outcomes linearly associated with Big Five personality traits?
A number of previous studies (Ozer & Benet-Martínez, 2006, Roberts, Kuncel, Shiner, Caspi & Goldberg, 2007 Soto, 2019) have investigated this research question and provide estimates for the expected associative effect sizes. We will not state these separately here but use the reported coefficients in (Soto, 2019) as a benchmark. We will only be able to investigate a limited number of associations, because not all variables from Soto, 2019 will be available in our dataset (see the section “Variables” for details) (H6).

Tertiary
Methodological effects
While machine learning methods have been successfully applied in many scientific disciplines, psychological science has been slower in adapting those methods (Stachl, Pargent, et al., 2020). One possible reason for this is that machine learning methods might not be able to unfold their full potential, because conventional psychological measurements are error-prone and low in reliability (Jacobucci & Grimm, 2020). We will investigate this question as part of this study by comparing the predictive performance of linear and non-linear models, when using self-reported personality traits as predictors (H4, H2).

Do lower level personality representations predict behavior better? Previous research has compared the predictiveness of personality items (nuances) vs domain scores for self-reported outcomes and abstract behaviors (Seeboth & Mõttus, 2018, Hall & Matz, 2020).
Results suggest that lower level measures (items) allow for better predictions of life outcomes.
We will re-investigate this research question as part of this study (H3).

Hypotheses

Please provide hypotheses for predicted results. If multiple hypotheses, uniquely number them (e.g. H1, H2a, H2b,) and refer to them the same way at other points in the registration document and in the manuscript.

H1 (life outcome prediction):
H1a: It will be possible to predict (out-of-sample) life outcomes with self-reported personality traits (all levels) above baseline.
H1b: It will be possible to predict (out-of-sample) life outcomes with automatically collected behavioral data (smartphone sensing), above baseline.

H2: Models using behavioral sensing data, will have a **higher** predictive performance than models using personality trait measures as predictors (average performance across all life outcomes).

H3: (levels of personality): Predictive performance for life-outcomes from personality trait measures, will **decrease** alongside the level of aggregation personality trait measures across models (from items/nuances (highest) to facets, to domain scores (lowest)).

H4: (shared method variance): Linear models (elastic net) using self-reported big five personality traits (all levels) will have a **higher** performance in the prediction of self-reported life outcomes than non-linear models with the same predictor variables (Random Forest).

H5: (multimodality, interactivity): Non-linear models, combining self-reports (personality traits) with behavioral sensing data will have the **highest** predictive performance overall (average performance across all life outcomes).

H6 (1:40) (replication): For available life outcomes, we assume significant correlations with personality traits, of equal association as reported in Soto, 2019 (see outcome measures in green font below, the table below and Table 1 in Soto, 2019).

Predicted trait-outcome associations							
Category	**Subcategory**	**Outcome: Sub-outcome**	**E**	**A**	**C**	**N**	**O**
Individual	Happiness	Subjective well-being: Life satisfaction	POS			NEG	
Individual	Happiness	Subjective well-being: Positive affect	POS			NEG	
Individual	Happiness	Subjective well-being: Negative affect (reversed)	POS			NEG	
Individual	Happiness	Subjective well-being: Happiness	POS			NEG	
Individual	Spirituality & virtues	Religious beliefs and behavior		POS	POS		
Individual	Health	Risky behavior: Activity (reversed)			NEG		
Individual	Health	Risky behavior: Excessive alcohol use			NEG		
Individual	Health	Risky behavior: Drug use			NEG		
Individual	Health	Risky behavior: Tobacco use			NEG		
Individual	Health	Coping	POS			NEG	
Individual	Health	Resilience	POS				
Individual	Psychopathology	Substance abuse			NEG		POS
Individual	Psychopathology	Depression	NEG			POS	
Interpersonal	Peer & family relations	Dating variety	POS				
Interpersonal	Peer & family relations	Attractiveness	POS				
Interpersonal	Romantic relations	Satisfaction	POS	POS	POS	NEG	
Social institut	Occupational choice &	Performance			POS		
Social institut	Occupational choice &	Satisfaction	POS			NEG	
Social institut	Occupational choice &	Commitment	POS			NEG	
Social institut	Occupational choice &	Job attainment		POS			
Social institut	Occupational choice &	Involvement	POS				
Social institut	Occupational choice &	Financial security				NEG	
Social institut	Political attitudes & val	Right-wing authoritarianism					NEG
Social institut	Political attitudes & val	Conservatism			POS		NEG
Social institut	Criminality	Antisocial behavior			NEG	POS	

This table was copy pasted from the OSF repository of Soto et al. 2019 and altered for the purpose of this pre-registration.

Variables

Which variables will be used? (see [Variables](#) in the basic protocol for an extensive overview of all available variables). This section shall be used to unambiguously clarify which variables are used to operationalize the specified hypotheses. Please (a) list all variables that will be used in this study and (b) explicitly state the functional role of each variable (i.e., independent variable, dependent variable, covariate, mediator, moderator). It is important to (c) specify for each hypothesis how it is operationalized, i.e., which variables will be used to test the respective hypothesis and how the hypothesis will be operationally defined in terms of these variables. This section is closely related to the statistical models used to test the hypotheses.

We will use the 300 items of the Big Five Structure Inventory (BFSI) and the aggregated latent trait scores on facet (30) and domain level (derived from partial credit model). While the full range of behavioral sensing variables that will be extracted from the raw sensing data is undefined at this

point, it is clear that we will use data from all sensing categories, as described in the basic pre-registration protocol of the panel study. In contrast to a top down, literature-driven modeling approach, we will apply a data-driven, bottom up approach to identify if the full breadth of mobile sensing variables will allow for the prediction of life outcomes at comparable performance to self-reported personality traits. It is neither intended nor possible to define all used behavioral variables at this point. However, all features included in the models will be held to the same strict evaluation criteria as the self-reported personality trait scores (cross-validated, out-of-sample prediction). We will include the following measures, also referenced in the basic protocol:

PERSONALITY TRAITS
Big Five (BFSI) (300 items)

LIFE OUTCOMES (inspired by Ozer & Benet-Martínez, 2006 and Soto, 2019)
IN Soto, 2019 (not same instruments),
ADDITIONAL LIFE OUTCOMES (measures only available in the present study, sensing-based outcomes might not be available for all participants, sensed outcomes (e.g., physical activity) will only be predicted by self-reported personality traits and sensing data that does not include the outcome measure that is predicted)
Aggregated scores will be computed by first, summing items of the same measure (i.e., items of a questionnaire) in an instance (while accounting for poling) and subsequently taking the mean across all instances (repeated measures of the same variable).
INDIVIDUAL OUTCOMES

- Subjective Wellbeing
    - Life satisfaction (satisfaction with life scale - SWLS, 5 items)
    - Positive affect (PANAS)
    - Negative Affect (PANAS)(reversed)
    - Mental wellbeing (WEMWBS)
    - Mean happiness (4 items, repeatedly measured (T = 6), aggregated score)
    - Vitality (2 items, per 4 weeks, aggregated score)
    - Flourishing (8 items, per 4 weeks, aggregated score)
- Religious believes (3 items, aggregated)
- Health
    - Physical activity - (sensing) (reversed)
    - Compulsive Smartphone usage (3 items, per 4 weeks, aggregated score)
    - Drug Use
        - Smoking cigarettes (1 item, per 4 weeks, aggregated score)
        - Alcohol (1 item, per 4 weeks, aggregated score)
        - Excessive alcohol use (1 item, per 4 weeks, aggregated score)
        - Coffee (1 item, per 4 weeks, aggregated score)
        - Illegal Drugs (1 item, per 4 weeks, aggregated score)
    - Coping (Brief COPE)
    - Dating Variety - proxy frequency dating app usage (sensing)
    - Attractiveness - proxy Dating App Notifications (sensing)
    - Mean Sleep Quality (PSQI, 19 items)
- Psychopathology
    - Depression (PHQ-9, 8 items - suicide item excluded)
    - Psychosis proneness (Disintegration, 20 items)

- Substance abuse (1 item, per 4 weeks, aggregated score)

INTERPERSONAL OUTCOMES

- Peer, family & romantic relations
  - Dating Variety - Proxy Dating/Mating App-Use frequency (sensing)
  - Attractiveness - Proxy Relation Dating Variety/Dating App Notifications (sensing)
  - Attachment style (ECR-R, 36 items)
  - Relationship satisfaction (CSI-4, Short couple satisfaction index, 4 items)

SOCIETAL/INSTITUTIONAL OUTCOMES

- Occupational Outcomes
  - Highest level of education (1 item)
  - Occupational status (1 item)
  - Occupational satisfaction (1 item)
  - Task performance (4 items, per 4 weeks, aggregated score)
  - Creative performance (3 items, per 4 weeks, aggregated score)
  - Role ambiguity (2 items, per 4 weeks, aggregated score)
  - Procrastination (3 items, per 4 weeks, aggregated score)
  - Job stress (3 items, per 4 weeks, aggregated score)
  - Perceived socio-economic status (1 item)
  - Income per month in € (1 item)
- Political attitudes and values
  - Political Interest (1 item)
  - Political Attitude (Conservatism) (3 items)
  - Authoritarianism (KSA-3)
- Environmental Attitudes and values (NEP scale, 15 items)
- Antisocial Personality (Dark Triad, 12 items)

Sampling Plan

More details on sampling for this study can be found in the general pre-registration protocol.

Existing Data
Registration prior to access to the data.

Explanation of existing data
Data collection is ongoing as described in the general registration protocol. I have (and will continue to) received updates on the sample size and the demographic composition of the sample. However, I did not have access to the data, up to the submission of this registration.

Data collection procedures
See the basic registration protocol for more details.

Sample size & rationale
The targeted initial sample size of the panel study will be 800. The sample size was chosen to maximize its representativeness under given financial constraints. After the initial recruitment period (till 27.05.2020, 24:00h CET), approximately 800 participants will have been recruited. The final sample size (at the end of the 6 month period) could be as low as 400 participants (based on estimates of the panel provider). While it is difficult to conduct a power analysis for the predictive modeling part of the pre-registration (cross-validated out of sample performance will be used), we provide the following estimates for two scenarios of participant dropout for the associative analyses (bivariate Pearson correlations). Conservative: Assuming a high drop out of participants from the study (400 remaining) we will have approximately 100% power to detect effect sizes of r = 0.5, 100% power for r = 0.3, and only 51% power for effects of sizes of r = 0.1, at alpha = 0.05. Smallest detectable effect size at 100% power is r = 0.3, at 80% power r = 0.14. Optimistic: Assuming a low drop out of participants from the study (600 remaining) we will have approximately 100% power to detect effect sizes of r = 0.5, 100% power for r = 0.3, and only 69% power for effects of sizes of r = 0.1 at alpha = 0.05. Smallest detectable effect size at 100% power is r = 0.25, at 80% power is r = 0.11.

Analysis Plan

Preprocessing

Inclusion criteria (e.g., criteria for including (1) participants (e.g., Do you only use a subsample?, (2) study days (e.g., only weekdays, certain number of study days), (3) any other criteria concerning data quality (e.g., only days with at least x% of logging data) etc. If you cannot specify these aspects now, please state why.
Data availability: We will exclude participants that have not provided data on personality traits, and participants with no data on the respective life outcome measure. If participants have provided data on a subset of life outcomes, we will only include them in the prediction models for the life outcomes they have provided data for. We will exclude participants with less than 30 days of mobile sensing data, in

total. Also, we will exclude participants that did not use apps on their phones.

Straightlining:
We will exclude respondents whose responses to the BFSI items have a within-person standard deviation of less than 0.4 (on a four-point response scale). This means that less than 25% of the item values deviate by 1 point on the Likert scale.

Speeding:
We will exclude respondents who completed the personality survey in less than one-third of the median completion time (as recorded by Unipark).

The data exclusion criteria with regard to straightlining and speeding were inspired by the preregistration protocol of Soto, 2019.

Definition of variables based on smartphone sensing. Please specify your degrees of freedom in variable extraction procedures, e.g.,

- *time information (e.g., what does night, daily, weekend exactly mean?)*
- *Aggregation measures (e.g., measures of central tendency/dispersion).*

If you cannot specify these aspects now, please state why.

Nighttime will be defined as the time between 20:00:01 and 06:00:00.
Weekends will start on Saturday at 00:00:00 and end on Sunday at 23:59:59.
Weektime will start on Monday at 00:00:00 and will end on Friday 23:59:59.

For variables, visually following a Gaussian distribution we will use the arithmetic mean and standard deviation as estimates for the central tendency and dispersion in the data. If outliers are present in the data (based on visual inspection) or data is not normally distributed we will use the Huber M estimator and the median absolute deviation (MAD) as estimates of central tendency and data dispersion.

Further preprocessing steps (e.g., transformation of data, handling of missing data/outliers etc.)

Data will be checked for plausibility and logging errors. This process will involve (but will not be limited to) sanity checks, such as if events could possibly have happened (e.g., activity while screen is off, screen off event without preceding screen on event), logging errors (e.g., cloned events). GPS data will be checked additionally for validity by excluding data from participants with GPS spoofing software installed.

Data preprocessing will be performed within each sub-fold of the resampling procedure (see Data Analysis for details). In that regard, missing data will be imputed with the median of the remaining values for elastic net models and with 2 times the maximum value for random forest models. For

outlier handling see the “Definition of variables” section above. Variables with no or little variation (i.e., constant values) will be removed during the nested preprocessing. Based on previous work (Stachl et al., 2019) we will use an arbitrary threshold of 2% unique values that will be applied during each resampling iteration. With as similar initial feature set as in (Stachl et al., 2019) we expect this threshold to effectively reduce the number of variables by around 90%.

As part of the pre-processing steps, we will possibly use dimensionality reduction methods (e.g., PCA, Autoencoders, Tensor Factorization) and or feature selection procedures (e.g., filter methods, forward search), to reduce the number of predictor variables in the models.

For elastic net models, the data will be centered and scaled as part of the modeling process - this is necessary to allow for effective regularization of beta-coefficients in the model.

Data Analyses

Statistical models
Please specify the statistical model (e.g. t-test, ANOVA, LMM) or algorithms that will be used to test each of your hypotheses. Give all necessary information about model specification (e.g., variables, interactions, planned contrasts) and follow-up analyses. Include model selection criteria (e.g., fit indices), corrections for multiple testing, and tests for statistical violations, if applicable. Please also indicate Inference Criteria (e.g., p-values, effect sizes, performance measures etc.).

PREDICTIVE ANALYSES (H1-H5)
We will use a supervised predictive modeling approach to answer the research questions in our study. We will use elastic net, regularized linear models (glmnet package) and random forest (ranger package) models.

For joint model optimization and evaluation we will use nested resampling. In the inner loop (5 or 10-fold CV) of the resampling approach we will perform data pre-processing and optimize hyperparameters. If the computational burden is too high we will not optimize model hyperparameters. In the outer resampling loop (10-times repeated 5-fold CV) we will evaluate the models performance on unseen test data. The model performance/generalization error will be calculated by aggregating across the resampling folds of the outer loop. We will provide standard deviations for the performance measure across the resampling folds of a respective model. For binary or multinomial (or ordinal) classification problems we will apply stratified sampling procedures to ensure representative and comparable class membership in all folds. In the case of unbalanced classification problems, we will apply up/oversampling, classification-threshold tuning, and/or use cost-sensitive classification approaches.

To test whether a model performs significantly better than the baseline model, we will fit baseline models which will constantly predict the mean (regression) or most prevalent group (classification)

as represented in the respective training data. To evaluate the predictiveness of our models and to test our hypotheses, we will compare all models with the performance of the respective baseline model. We will consider models as predictive if the median performance in R^2 is greater than in the comparable baseline or comparison model and is greater than $R^2 = 0.01$ in absolute terms. We will report at least two performance measures for regression and classification models. For classification models we will at least report the F1 and AUC, the true positive rate, and the true negative rate. For regression tasks we will report the R^2 metric, the mean absolute error (MAE) and the Spearman sample correlation.

ASSOCIATIVE ANALYSIS (H6 1:40)
In addition to the prediction models we will compute pairwise Pearson correlation coefficients and 95% bootstrapped confidence intervals between the Big Five personality trait measures (domain level) and the life outcome measures. As a mental exercise, we will use the following 3 methodological approaches to test H6 1:40. We will specifically report if and which criteria are met with regard to replication of these effects:

1. Significance
   Under this criterion we will declare the associative effects between life outcomes and personality traits as replicated if the replication effect is significant and shows the same direction of association. Consistent with Soto et al., 2019, will use an alpha level of 0.05.

2. Effect size
   Under this criterion we will declare the associative effects between life outcomes and personality traits as replicated if the replication effect is located within the prediction interval of the original effect (in Soto et al., 2019).

3. Sceptical p-value
   Under this criterion we will calculate sceptical p-values (Held, 2020) as a criterion for determining replication success. It is dependent on the finally achieved sample size whether previous findings will be replicable under this, most stringent condition.

For all three approaches we will use the ReplicationSuccess R package in R.

Predicted trait-outcome associations								
Category	**Subcategory**	**Outcome: Sub-outcome**	**E**	**A**	**C**	**N**	**O**	**Analysis**
Individual	Happiness	Subjective well-being: Life satisfaction	POS			NEG		Pearson-Product Moment Correlation
Individual	Happiness	Subjective well-being: Positive affect	POS			NEG		Pearson-Product Moment Correlation
Individual	Happiness	Subjective well-being: Negative affect (reversed)	POS			NEG		Pearson-Product Moment Correlation
Individual	Happiness	Subjective well-being: Happiness	POS			NEG		Pearson-Product Moment Correlation
Individual	Spirituality & virtues	Religious beliefs and behavior		POS	POS			Pearson-Product Moment Correlation
Individual	Health	Risky behavior: Activity (reversed)			NEG			Pearson-Product Moment Correlation
Individual	Health	Risky behavior: Excessive alcohol use			NEG			OLS Regression on C, age, sex, and race/ethnicity
Individual	Health	Risky behavior: Drug use			NEG			OLS Regression on C, age, sex, and race/ethnicity
Individual	Health	Risky behavior: Tobacco use			NEG			OLS Regression on C, age, sex, and race/ethnicity
Individual	Health	Coping	POS			NEG		Partial correlation, controlling for age
Individual	Health	Resilience	POS					OLS Regression on E, N, age, and sex
Individual	Psychopathology	Substance abuse			NEG		POS	Partial correlation, controlling for sex
Individual	Psychopathology	Depression	NEG			POS		Partial correlation, controlling for sex
Interpersonal	Peer & family relations	Dating variety	POS					Partial correlation, controlling for sex
Interpersonal	Peer & family relations	Attractiveness	POS					Partial correlation, controlling for sex
Interpersonal	Romantic relations	Satisfaction	POS	POS	POS	NEG		Partial correlation, controlling for sex, separately by relationship status
Social institut	Occupational choice &	Performance			POS			Partial correlation, controlling for age and sex
Social institut	Occupational choice &	Satisfaction	POS			NEG		Pearson-Product Moment Correlation
Social institut	Occupational choice &	Commitment	POS			NEG		Pearson-Product Moment Correlation
Social institut	Occupational choice &	Job attainment		POS				Partial correlation, controlling for age
Social institut	Occupational choice &	Involvement	POS					Partial correlation, controlling for age
Social institut	Occupational choice &	Financial security				NEG		Partial correlation, controlling for age
Social institut	Political attitudes & val	Right-wing authoritarianism					NEG	Pearson-Product Moment Correlation
Social institut	Political attitudes & val	Conservatism			POS		NEG	Pearson-Product Moment Correlation
Social institut	Criminality	Antisocial behavior			NEG	POS		Pearson-Product Moment Correlation

Planned exploratory analysis (Optional)

We will use interpretable machine learning techniques to investigate which sensed behaviors, groups of behaviors, and personality items/facets/domains were most predictive for each life outcome measure.

We will explore pairwise Pearson correlations between additional life outcomes (see blue life outcomes in the Variables section of this registration).

References

Bouckaert, R. R., & Frank, E. (2004). Evaluating the replicability of significance tests for comparing learning algorithms. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3056*, 3–12. https://doi.org/10.1007/978-3-540-24775-3_3

Boyd, R. L., Pasca, P., & Lanning, K. (2020). The Personality Panorama: Conceptualizing Personality Through Big Behavioural Data. *European Journal of Personality*. https://doi.org/10.1002/per.2254

Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: a new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, Vol. 18, pp. 63–68. https://doi.org/10.1016/j.cobeha.2017.07.017

Ellis, D. A., Davidson, B. I., Shaw, H., & Geyer, K. (2019). Do smartphone usage scales predict behavior? International Journal of Human-Computer Studies, 130, 86–92. https://doi.org/10.1016/j.ijhcs.2019.05.004.

Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, *23*(5), 369–401. https://doi.org/10.1002/per.724

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, *74*(5), 1337–1349. https://doi.org/10.1037/0022-3514.74.5.1337.

Hall, A. N., & Matz, S. C. (2020). Targeting Item-level Nuances Leads to Small but Robust Improvements in Personality Prediction from Digital Footprints. *European Journal of Personality*, per.2253. https://doi.org/10.1002/per.2253

Harari, G. M., Gosling, S. D., Wang, R., & Campbell, A. T. (2015). Capturing Situational Information with Smartphones and Mobile Sensing Methods. *European Journal of Personality*, *29*(5), 509–511. https://doi.org/10.1002/per.2032

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, *11*(6), 838–854. https://doi.org/10.1177/1745691616650285

Harari, G. M., Müller, S. R., Stachl, C., Wang, R., Wang, W., Bühner, M., … Gosling, S. D. (2019). Sensing Sociability: Individual Differences in Young Adults' Conversation, Calling, Texting, and App Use Behaviors in Daily Life. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000245

Held, L. (2020). A new standard for the analysis and design of replication studies. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *183*(2), 431–448. https://doi.org/10.1111/rssa.12493

Jacobucci, R., & Grimm, K. J. (2020). Machine Learning and Psychological Research: The Unexplored Effect of Measurement. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *15*(3), 809–816. https://doi.org/10.1177/1745691620902467

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the Prediction of Consequential Outcomes. *Annual Review of Psychology*, *57*(1), 401–421.

https://doi.org/10.1146/annurev.psych.57.102904.190127

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *2*(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Seeboth, A., & Mõttus, R. (2018). Successful Explanations Start with Accurate Descriptions: Questionnaire Items as Personality Markers for More Accurate Predictions. *European Journal of Personality*, *32*(3), 186–201. https://doi.org/10.1002/per.2147

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis. *Cyberpsychology, Behavior, and Social Networking*, *21*(4), 217–228. https://doi.org/10.1089/cyber.2017.0384

Soto, C. J. (2019). How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project. *Psychological Science*, *30*(5), 711–727. https://doi.org/10.1177/0956797619831612

Soto, C. J. (in press). Do links between personality and life outcomes generalize? Testing the robustness of trait-outcome associations across gender, age, ethnicity, and analytic approaches. *Social Psychological and Personality Science*

Stachl, C., Au, Q., Schoedel, R., Buschek, D., Völkel, S., Schuwerk, T., … Bühner, M. (2019, June 12). Behavioral Patterns in Smartphone Usage Predict Big Five Personality Traits. https://doi.org/10.31234/osf.io/ks4vd

Stachl, C., Hilbert, S., Au, J. Q., Buschek, D., De Luca, A., Bischl, B., … Bühner, M. (2017). Personality Traits Predict Smartphone Usage. European Journal of Personality, 31(6), 701–722. https://doi.org/10.1002/per.2113

Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S., … Bühner, M. (2020). Personality Research and Assessment in the Era of Machine Learning. European Journal of Personality, per.2257. https://doi.org/10.1002/per.2257

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger(Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). Retrieved from http://books.google.com/books?hl=en&lr=&id=XHwS3PU6uroC&oi=fnd&pg=PA224&dq=The+self-report+method&ots=JBJjUXTRTG&sig=_i_Lja4kJJxGgW5PkdHfdhM63Tw