Preregistration for Quantitative Research in Psychology (PRP-QUANT) Template

Title

T1 Title
The title should be focused and descriptive, using relevant key terms to reflect what will be done in the study. Use title case (https://apastyle.apa.org/style-grammar-guidelines/capitalization/title-case).
Pretesting Video Speeches for an Experiment on the Effects of Two Dimensions of Servant Leadership on Performance and Examining the Strength of Potential Instruments

T2 Contributors, Affiliations, and Persistent IDs (recommend ORCID iD)
Provide in separate entries the full name of each contributor, each contributor's professional affiliation, and each contributor's persistent ID. See ORCID iD for an example of persistent ID (https://orcid.org/). Optional: include the intended contribution of each person listed (e.g. statistical analysis, data collection; see CRediT, https://casrai.org/credit/).
Annika F. Schowalter, University of Bamberg, https://orcid.org/0000-0001-9478-7438 Judith Volmer, University of Bamberg, https://orcid.org/0000-0003-4476-6537

T3 Date of Preregistration
This is assigned by the system upon preregistration submission.

T4 Versioning information

This is assigned by the system upon submission of original and subsequent revisions. Should be a persistent identifier, if not a DOI.

T5 Identifier

This unique identifier is assigned by the system upon submission.

T6 Estimated duration of project

Include best estimate for how long the project will take from preregistration submission to project completion.

The pretest is part of a resubmission at *The Leadership Quarterly* (Special Issue Call: *Beyond the ritualized use of questionnaires: Toward a science of actual behaviors and psychological states*).

T7 IRB Status (Institutional Review Board/Independent Ethics Committee/Ethical Review Board/Research Ethics Board)

If the study will include human or animal subjects, provide a brief overview of plans for the treatment of those subjects in accordance with established ethical guidelines. If appropriate institutional approval has been obtained for the study, provide the relevant identifier here. If the study will be exempt from ethical board review, provide reasoning here.

The experiment was approved by the local Ethical Review Board (University of Bamberg).

T8 Conflict of Interest Statement

Identify any real or perceived conflicts of interest with this study execution. For example, any interests or activities that might be seen as influencing the research (e.g., financial interests in a test or procedure, funding by pharmaceutical companies for research).

We have no conflicts of interest to declare.

T9 Keywords

Include terms specific to your topic, methodology, and population. Use natural language and avoid words used in the title or overly general terms. If you need help with keywords, try a keyword search using your proposed keywords in a search engine to check results.

Stewardship, authenticity, instrument strength

T10 Data accessibility statement and planned repository

"We plan to make the data available (yes / no)
If "yes", please specify the planned data availability level by selecting one of the options:

- Data access via download; usage of data for all purposes (public use file)
- Data access via download; usage of data restricted to scientific purposes (scientific use file)
- Data access via download; usage of data has to be agreed and defined on an individual case basis
- Data access via secure data center (no download, usage/analysis only in a secure data center)
- Data available upon email request by member of scientific community
- Other (please specify)

Yes.
Data access via download; usage of data for all purposes (public use file)

T11 Optional: Code availability

We plan to make the code available (yes / no).
If "yes", please specify the planned code availability level (use same descriptors of data in T10).

Yes.
Code available upon email request by member of scientific community

T12 Optional: Standard lab practices

Standard lab practices refer to a (timestamped) document, software package, or similar, which specifies standard pipelines, analytical decisions, etc. which always apply to certain types of research in a lab. Specify here and refer to at the appropriate positions in the remainder of the template:
We plan to make the standard lab practices available (yes / no).
If "yes", please specify the planned standard lab practices availability level (use same descriptors of data in T10).

Yes.
Data and procedure available upon email request by member of scientific community

Abstract

(150 words)

A1 Background

(See introduction I1)

The concept of servant leadership has received growing interest in recent years. To date, however, there are no clean causal investigations of the effects of servant leadership.

A2 Objectives and Research questions

(See introduction I2)

We are the first to use an instrumental variable approach in the context of servant leadership and created new material based on the literature. This pre-study thus serves to test how followers perceive the leadership manipulation and to examine the strength of our potential instrumental variables.

A3 Participants

(See methods M4)

We aim to recruit at least 100 working-age individuals using a common panel provider.

A4 Study method

(See methods M10-14)

Our hypotheses will be tested by means of an experimental online study with two groups (authenticity and stewardship vs. neutral speech).

Introduction

(no word limit)

I1 Theoretical background

Provide a brief overview that justifies the research hypotheses.

The theoretical background is mainly identical to the following preregistration by Schowalter and Volmer (2021). Thus, we refer the reader to the publication for in-depth information and describe only the changes in this preregistration.

In the pretest based on Schowalter and Volmer (2021), we found no relevant differences between the perceptions of the female and the male actor. The gender of the actors was thus not an appropriate instrumental variable. Likewise, we discarded agreeableness as an instrumental variable as its relations with perceptions of servant leadership (i.e., authenticity and stewardship) and charismatic leadership were equally strong. However, we will still include agreeableness to test it again. If it is not appropriate as an instrumental variable, it could be an important control variable.
Confirming our hypothesis, the leadership manipulation affected servant leadership perceptions. However, it was not sufficiently strong as an instrument. Thus, we will improve the speeches and conduct the pre-study again; without varying leader gender.

If the relevance of the instruments were not given, we will only calculate the main effect of the leadership manipulation on performance in the planned experiment for which we will use the pre-study results.

I2 Objectives and Research question(s)

Outline objectives and research questions that inform the methodology and analyses (below).

We aim to pretest our manipulation. Moreover, we will derive estimates for the power analysis to determine the appropriate sample size and calculate the duration of the cognitive performance task for the planned experiment based on the data.

Research questions:

1. Do individuals in the authenticity and stewardship condition perceive the leader as higher on the two dimensions than individuals in the neutral condition?
2. Is the manipulation as an instrumental variable strong enough for the planned analyses?
3. Is agreeableness an appropriate instrumental variable?

I3 Hypothesis (H1, H2, …)

Provide hypothesis 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: Individuals exposed to the authenticity and stewardship condition will perceive the leader as more servant than individuals exposed to the neutral speech.
H2: The manipulation of the leadership style (authenticity and stewardship vs. neutral) is a relevant instrumental variable to instrument followers' perceptions of authenticity and stewardship.
H3: Participants' agreeableness is a relevant instrumental variable to instrument followers' perceptions of leadership on task performance.

I4 Exploratory research questions (if applicable; E1, E2, ...)

If planning exploratory analyses, provide rationale for them here. If multiple exploratory analyses, uniquely number them (E1, E2, ...) and refer to them in the same way in the registration document and in future publications.

As in Schowalter and Volmer (2021), we plan to investigate whether our leadership manipulation also influences perceptions of charismatic leadership. If there are no differences in perceptions of charismatic leadership between the two conditions (stewardship & authenticity vs. neutral), this will strengthen our argument regarding the exclusion restriction.

Method

M1 Time point of registration

Select one of the options:

- Registration prior to creation of data
- Registration prior to any human observation of the data
- Registration prior to accessing the data
- Registration prior to analysis of the data
- Other (please specify; might include if T1 longitudinal data has been analyzed, but T2 has not yet been analyzed)

Registration prior to creation of data

M2 Proposal: Use of pre-existing data (re-analysis or secondary data analysis)

Will pre-existing data be used in the planned study? If yes, indicate if the data were previously published and specify the source of the data (e.g., DOI or APA style reference of original publication). Specify your level of knowledge of the data (e.g., descriptive statistics from previous publications), whether or not this is relevant for the hypotheses of the present study, and how it is assured that you are unaware of results or statistical patterns in the data of relevance to the present hypotheses.

Sampling Procedure and Data Collection

M3 Sample size, power and precision

(1) Relevant sample sizes: e.g., single groups, multiple groups, and sample sizes (or sample ranges) found at each level of multilevel data. (2) Provide power analysis (e.g. power curves) for fixed-N designs. For sequential designs, indicate your 'stopping rule' such as the points at which you intend to be viewing your data and in any way analyzing them (e.g., t-tests and correlations, but even descriptively such as with histograms).

As there are no previous data we could rely on to derive a reasonable estimate of the needed sample size through power analysis, and because the preliminary study is intended to serve as a basis to calculate the sample size for the planned experiment, we follow recommendations in the literature. In total, we aim to recruit at least 100 individuals in accordance with Lonati et al. (2018), who recommend at least 50 individuals per condition in experimental studies.

M4 Participant recruitment, selection, and compensation

Indicate (a) methods of recruitment (e.g., subject pool advertisement, community events, crowdsourcing platforms, snowball sampling); (b) selection and inclusion/exclusion criteria (e.g., age, visual acuity, language facility); (c) details of any stratification sampling used; (d) planned participant characteristics (gender, race/ethnicity, sexual orientation and gender identity, SES, education level, age, disability or health status, geographic location); (e) compensation amount and method (e.g., same payment to all, pay based on performance, lottery).

We plan to use Respondi, a common panel provider, with a UK sample. All participants should be English-speaking. To obtain a sample that is as representative as possible, we plan to use quotas reproducing the distribution of age and gender in the working-age population in the UK. The numbers are derived from the latest Census. Using the same sample characteristics in the pretest as in the planned experiment, we want the results to

be as comparable as possible to the experiment (where the sample should be representative).
Males: *N* = 49
Females: *N* = 51
Age 18-24 years: *N* = 15
Age 25-29 years: *N* = 11
Age 30-39 years: *N* = 22
Age 40-49 years: *N* = 20
Age 50-59 years: *N* = 22
Age 60-64 years: *N* = 10
Participants will receive the usual compensation from the panel provider. Additionally, the research team will donate 0.03 GBP to the charity World Vision (a non-profit relief, development, and advocacy organization) for each letter that is correctly decoded in the course of the performance task.

M5 How will participant drop-out be handled?

Indicate any special treatment for participants who drop out (e.g., there is follow-up in a manner different from the main sample, last value carried forward) or whether participants are replaced.

We will use only one measurement point.

M6 Masking of participants and researchers

Indicate all forms of masking and/or allocation concealment (e.g., administrators, data collectors, raters, confederates are unaware of the condition to which participants were assigned).

It is an online experiment with a randomized assignment to groups and a manipulation via recorded videos; all researchers will be unaware of the condition to which participants will be assigned.

M7 Data cleaning and screening

Indicate all steps related to data quality control, e.g., outlier treatment, identification of missing data, checks for normality, etc.

Cases with incorrect answers on at least one of the three attention/comprehension check items will be excluded already during data collection (quality fail). We will plot the data (histograms and box-whisker diagrams) and check the distribution of the servant leadership, the charismatic leadership, and agreeableness scores as well as the descriptive data for plausibility.
We will calculate the reliabilities and the means of the servant leadership, the charismatic leadership, and the agreeableness items and check the values for normality.

M8 How will missing data be handled?

Indicate any procedures that will be applied during the analysis to deal with missing data, such as (a) case deletions; (b) averaging across scale items (to handle missing items for some); (c) test of missingness (MAR, MCAR, MNAR assumptions; (d) imputation procedures (FIML vs. MI); (e) Intention to treat analysis and per protocol analysis (as appropriate).

Cases with no data on the variables will be deleted.

M9 Other information (optional)

For example, training of raters/participants or anything else not yet specified.

Conditions and design

M10 Type of study and study design

Indicate the type of study (e.g., experimental, observational, crosssectional vs. longitudinal, single case, clinical trial) and planned study design (e.g., between vs. within subjects, factorial, repeated measures, etc.), number of factors and factor levels, etc..

We will use an experimental online study with two conditions (authenticity & stewardship vs. neutral speech).

M11 Randomization of participants and/or experimental materials

If applicable, describe how participants are assigned to conditions or treatments, how stimuli are assigned to conditions, and how presentation of tests, trials, etc. is randomized. Indicate the randomization technique and whether constraints were applied (pseudo-randomization). Indicate any type of balancing across participants (e.g., assignments of responses to hands, etc.).

Participants will be randomly assigned to groups by a randomization trigger that is linked with filters. The filters allow for showing only one video to each participant.
The items to measure servant leadership, charismatic leadership, and agreeableness as well as the sequence of the lines of code (performance task) will be randomized.

M12 Measured variables, manipulated variables, covariates

This section shall be used to unambiguously clarify which variables are used to operationalize the hypotheses specified above (item I3). Please (a) list all measured variables, 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. The description here shall be consistent with the statistical analysis plans specified under AP6 (below).

Manipulated variable (two videos):

- Leadership style
    - The manipulation consists of two factors: stewardship and authenticity vs. neutral.
    - The manipulation serves as the independent variable for Hypothesis 1. Additionally, we will examine its strength as an instrumental variable (Hypothesis 2).

Measured variables:

- Perceptions of stewardship and authenticity
    - The variable serves as an outcome to test the effect of the leadership speeches (Hypothesis 1).
    - Moreover, it is used as the dependent variable in the first stage of the two-stage least squares (2SLS) procedure to test its relevance as an instrumental variable.
- Perceptions of charismatic leadership
    - We will test whether differences in charismatic leadership perceptions occur depending on the leadership manipulation.
- Agreeableness
    - We will test the strength of agreeableness as an instrumental variable to instrument authenticity and stewardship perceptions.
- Task performance
    - Decoding task
    - We use a short form (seven minutes) of the experimental task to avoid deception.

- We will use it as a reference to decide on the duration of the task in the planned experiment.

Demographic & control variables

- Participants' age and gender, education
  - Participants' age and gender are necessary for quota sampling to approach representativeness of the sample regarding the UK working-age population (18-64 years).
  - We will use age and gender for randomization checks and include them in our analyses as control variables given that there is no full randomization.
- Additionally, we will include one instructed response item and two questions regarding the content of the study and leadership video as comprehension checks. If at least one of the three questions is answered incorrectly, we will exclude the respective participant (DeSimone & Harms, 2018).

Operationalization of hypotheses

- Hypothesis 1: β_1 should be significant when servant leadership perceptions are regressed on leadership style.
- Hypothesis 2 and 3: The *F*-statistic of leadership style exceeds the critical values by Stock and Yogo (2005) for the relevance of the instrument.

M13 Study Materials

Please describe any relevant study materials. This could include, for example, stimulus materials used for experiments, questionnaires used for rating studies, training protocols for intervention studies, etc.

Leadership Manipulation

As in the first pretest, participants will be randomly assigned to one of two videos (i.e., servant or neutral leader condition, played by an actor) of equivalent length.
Unlike the first pretest, we will not describe World Vision's fields of activity more factually and will not add any information about the organization that is not included in the servant leadership speech.
Before the two speeches will be recorded by the actor, we will conduct an objective manipulation check as in Schowalter and Volmer (2021). Two independent coders, unaware of the purpose of the study, will code both speeches sentence by sentence for the presence of the nine charismatic leadership tactics (Antonakis et al., 2021).

Questionnaires

After the video, participants will rate the four items to measure authenticity and three items to measure stewardship from the Servant Leadership Scale (Van Dierendonck & Nuijten, 2011). We adapted the items to fit the experimental context (i.e., instead of *leader*, we will use the name of the actor; we deleted *often*). An example item for authenticity is "[Name] shows his/her true feelings to the participants" and for stewardship "[Name] emphasizes the importance of focusing on the good of the whole.". The items will be rated on six-point Likert scales from 1 = *strongly disagree* to 6 = *strongly agree*.
Perceptions of charismatic leadership will be measured as in Antonakis et al. (2021) using the idealized influence and inspirational motivation subscales of the Multifactor Leadership Questionnaire (MLQ, Bass & Avolio, 1995). The 12 items will be rated on five-point Likert scales from 1 = *not at all* to 5 = *frequently, if not always*. An example item is "Talks optimistically about the future".

To measure agreeableness, participants will rate ten items from the 50-item International Personality Item Pool (IPIP: Goldberg, 1999) on a six-point Likert scale from 1 = *strongly disagree* to 6 = *strongly agree*. An example item is “am interested in people”.
Regarding the demographic variables, we will ask for participants’ age, gender, and education. Additionally, we will add one instructed response item (see Meade & Craig, 2012) and two questions regarding the content of the study and leadership video to make sure that participants have been attentive.

Cognitive Task
We will use Meslec et al.’s (2020) task as the performance task, where participants are asked to decrypt codes into meaningful English phrases. We used this task also in Schowalter and Volmer (2021). However, we will only use three of the original decoding schemes as two of them were not appropriate for the online version of the task (many people did not succeed in decrypting codes using the original schemes 2 and 4). We changed the codes accordingly, improved the display of the codes encrypted with scheme 3, and replaced “Trbuljak” with his first name.

M14 Study Procedures

Please describe here any relevant information about how the study will be conducted, e.g., the number and timing of measurement time points for longitudinal research, the number of blocks or runs per session of an experiment, laboratory setting, the group size in group testing, the number of training sessions in interventional studies, questionnaire administration for online assessments, etc.

The procedure is mainly the same as in Schowalter & Volmer (2021). However, we will only use two videos and collect data on the authenticity and stewardship subscales of servant leadership perceptions.

M15 Other information (optional)

Analysis plan

(NOTE: If this varies by hypothesis, repeat analysis plan for each)

AP1 Criteria for post-data collection exclusion of participants, if any
Describe all criteria that will lead to the exclusion of a participant's data (e.g. performance criteria, non-responding in physiological measures, incomplete data). Be as specific as possible.
Cases with no data on the variables will be deleted. Cases with implausible values will be excluded based on a multiple hurdle approach (Arthur et al., 2021; Goldammer et al., 2020).

AP2 Criteria for post-data collection exclusions on trial level (if applicable)
Describe all criteria that will lead to the exclusion of a trial or item (e.g. statistical outliers, response time criteria). Be as specific as possible.

AP3 Data preprocessing
Describe all data manipulations that are performed in preparation of the main analyses, e.g. calculation of variables or scales, recoding, any data transformations, preprocessing steps for imaging or physiological data (or refer to publicly accessible standard lab procedure, cf. T12).
We will calculate the means of the servant leadership, charismatic leadership, and agreeableness scale.

AP4 Reliability analysis (if applicable)
Specify the type of scale reliability that will be estimated, whether it is internal consistency (e.g. Cronbach's alpha, omega), test-retest reliability, or some other form (e.g., a

confirmatory factor analysis incorporating multiple factors as sources of variance). In a study involving measure development, researchers should specify criteria for removing items from measures a priori (e.g., largest factor loading magnitude, smallest drop in alpha-if-item removed).

We will calculate McDonald's omega for the servant leadership, charismatic leadership, and the agreeableness scale.

AP5 Descriptive statistics

Specify which descriptive statistics will be calculated for which variables. If appropriate, specify which indices of effect size will be used. If descriptive statistics are linked to specific hypotheses, explicitly link the information given here to the respective hypothesis.

We will calculate means and standard deviations of the servant leadership scale, the charismatic leadership scale, agreeableness, and age.

AP6 Statistical models (provide for each hypothesis if varies)

Specify the statistical model (e.g. t test, ANOVA, LMM) 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. Wherever unclear, describe how effect sizes will be calculated (e.g., for d-values, use the control SD or the pooled SD).

To check the randomized group assignment, we will regress participants' age, gender, and agreeableness on the conditions. If randomization was not successful regarding one or more of these variables, we will include the affected variables as control variables.
To test Hypothesis 1, we will use OLS regression as well.
To be relevant, the instrument has to significantly correlate with the servant leadership (i.e., authenticity and stewardship) scale ($Cov(z,x) \neq 0$). H2 and H3 can therefore be accepted if the first-stage *F*-statistics of the two-stage least squares (2SLS) procedure exceed the critical values by Stock and Yogo (2005).

AP7 Inference criteria

Specify the criteria used for inferences (e.g., p values, Bayes factors, effect size measures) and the thresholds for accepting or rejecting your hypotheses. If possible, define a smallest effect size of interest. If inference criteria differ between hypotheses, specify separately for each hypothesis and respective statistical model by explicitly referring to the numbers of the hypotheses. Describe which effect size measures will be

reported and how they are calculated.

Regarding the regression analysis (H1), the *p*-Value of the coefficient should be below the significance level of α = .05 and we will report the R^2.
The critical values for the first-stage *F*-statistics are displayed in Table 5.2 in Stock & Yogo (2005). Thus, the values should not be below *F* = 8.96 (H2 & H3).

AP8 Exploratory analysis (optional)

Describe any exploratory analyses to be conducted with your data. Include here any planned analyses that are not confirmatory in the sense of being a direct test of one of the specified hypotheses.

We will test whether there are differences between the neutral and the servant leadership (i.e., authenticity and stewardship) speech regarding charismatic leadership perceptions using OLS.

AP9 Other information (optional)

Other information optional

(NOTE: If needed, multiple lines with other information can be included)

O1 Other information (optional)
If there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related work such as replications or work that uses the same data, or other context that will be helpful for future readers would be appropriate here.
This preregistration is based on Schowalter and Volmer (2021) and describes mainly the changes. In-depth information can be found in the cited preregistration.

References

R1 References

Enter your references below. Use a consistent format (e.g., https://apastyle.apa.org/style-grammar-guidelines/references/examples)

Antonakis, J., d'Adda, G., Weber, R., & Zehnder, C. (2021). Just words? Just speeches? On the economic value of charismatic leadership. *Management Science*. Advance online publication. https://doi.org/10.1287/mnsc.2021.4219

Arthur, W., Hagen, E., & George, F. (2021). The lazy or dishonest respondent: detection and prevention. *Annual Review of Organizational Psychology and Organizational Behavior*, 8(1). https://doi.org/10.1146/annurev-orgpsych-012420-055324

Bass, B. M., & Avolio, B. J. (1995). *MLQ Multifactor leadership questionnaire for research: Permission set.* Mindgarden.

DeSimone, J. A., & Harms, P. D. (2018). Dirty data: The effects of screening respondents who provide low-quality data in survey research. *Journal of Business and Psychology*, 33(5), 559–577. https://doi.org/10.1007/s10869-017-9514-9

Goldammer, P., Annen, H., Stöckli, P. L., & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. *The Leadership Quarterly*, 101384. https://doi.org/10.1016/j.leaqua.2020.101384

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* (pp. 7–28). Tilburg University Press.

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management*, *64*(1), 19–40. https://doi.org/10.1016/j.jom.2018.10.003

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Meslec, N., Curseu, P. L., Fodor, O. C., & Kenda, R. (2020). Effects of charismatic leadership and rewards on individual performance. *The Leadership Quarterly*, *31*(6). https://doi.org/10.1016/j.leaqua.2020.101423

Schowalter, A. F., & Volmer, J. (2021). Pretesting video speeches to manipulate perceptions and examining strength of instruments for an experiment on servant leadership and performance. https://doi.org/10.23668/PSYCHARCHIVES.5146

Van Dierendonck, D., & Nuijten, I. (2011). The servant leadership survey: Development and validation of a multidimensional measure. *Journal of Business and Psychology*, *26*(3), 249–267. https://doi.org/10.1007/s10869-010-9194-1