Preregistration

(Re)Building Trust? Investigating the effects of open science badges on perceived trustworthiness of journal articles.

Jürgen Schneider[1], Samuel Merk[1], Tom Rosman[2], Augustin Kelava[1]

[1] University of Tübingen

[2] Leibniz Institute for Psychology Information

29. January 2020

Study Information

Title	(Re)Building Trust. Investigating the effects of open science badges on perceived trustworthiness of journal articles.
Description	The Replication Crisis diminishes trust in empirical sciences and with it the perceived value of science (Lupia, 2018). Open Science Practices (i.a. open data, open analysis script, open materials) are an increasingly popular approach to deal with challenges in replication and to rebuilt trust (Geukes, Schönbrodt, Utesch, Geukes, & Back, 2016). First investigations could, however, deliver no evidence toward the effect of Open Science Practices (OSP) on trustworthiness (Wingen, Berkessel, & Englich, 2019). However, this study investigated the effect on a discipline level (psychology)

with an abstract description of OSP. We want to shift the focus from discipline level to concrete individual journal articles and consider epistemic beliefs of readers to play a role (Merk & Rosman, 2018): Will visible OSP (vs. not visible vs. visibly non-OSP) foster perceived trustworthiness when reading journal articles of empirical studies? Hence we formulated the following research question:Will multiplistic epistemic beliefs moderate the relationship between OSP and trustworthiness?

Hypotheses

1. Confirmatory, H1: Visible OSP (vs. not visible vs. visibly non-OSP) influence the perceived trustworthiness (subscale integrity). Our assumption: The more openness, the more trustworthy with small to moderate effects: $\mu_1 < \mu_2 < \mu_3$. With the `bain` (Gu, Hoijtink, Mulder, & Lissa, 2019) package we will evaluate the following informative hypotheses using Bayes factors:
    1. $\mu_1 < \mu_2 < \mu_3$
    2. $\mu_1 = \mu_2 = \mu_3$
    3. $\mu_1 < \mu_2 = \mu_3$
    4. μ_1, μ_2, μ_3
2. Confirmatory, H2: The higher the topic specific multiplism, the lower the perceived trustworthiness (subscale integrity). Negative correlation.
3. Exploratory, H3: Topic specific multiplism moderates the effect of OSP on perceived trustworthiness (subscale integrity).
4. Exploratory, H4: Visible OSP (vs. not visible vs. visibly non-OSP) have a negative effect on topic specific multiplism.

Design Plan

Study type *Wording taken from OSF preregistration forms, since they are closed questions:*

Experiment. A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding	*Wording taken from OSF preregistration forms, since they are closed questions:* • For studies that involve human subjects, they will not know the treatment group to which they have been assigned. • Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments.
Study design	The design will include three conditions: visible Open Science Practices (visOSP), Practices not visible (nonvis) and visible non-Open Science Practices (nonOSP). Two of the (three) conditions are randomly chosen and randomized in their order within person. Realizing all three conditions within person would highlight the variation between conditions as too obvious and thus undermine blinding of subjects. visOSP condition: Subjects receive a title page of an empirical study (Title, Abstract, Keywords, Introduction, . . .) together with three Open Science badges. The badges are explained using hints in style of speech bubbles and indicate that the authors engaged in the OSP open data, open analysis script and open materials. nonvis condition: Subjects receive a title page of an empirical study (Title, Abstract, Keywords, Introduction, . . .) with no further information on Open Science, reflecting a "standard" journal article. nonOSP condition: Subjects receive a title page of an empirical study (Title, Abstract, Keywords, Introduction, . . .) together with three Open Science badges. The badges are explained using hints in style of speech bubbles and indicate that the authors did not engage in the OSP open data, open analysis script and open materials. As participants are exposed to more than one condition, we create all three conditions for three different empirical studies (topics). In doing so, we avoid participants to see one study topic twice under different conditions, which would undermine the blinding.
Randomization	• Randomization 1: Two of the three conditions will be randomly assigned to the participants. • Randomization 2: The order of presentation will be randomized between the two conditions, within the participant.

- Randomization 3: Within each of the six combinations of randomization 1 & 2, we will randomize the order of the topic between (topic 1-2, 2-3, 3-1).

Sampling Plan

Existing data

Wording taken from OSF preregistration forms, since they are closed questions: **Registration prior to analysis of the data**. As of the date of submission, the data exist and you have accessed it, though no analysis has been conducted related to the research plan (including calculation of summary statistics). A common situation for this scenario when a large data set exists that is used for many different studies over time, or when a data set is randomly split into a sample for exploratory analyses, and the other section of data is reserved for later confirmatory data analysis.

Explanation of existing data

We have already preregistered the study on https://osf.io/2zypf **prior to data collection**. There we planned two t-tests and hence used the according Bayes factor analysis. However, meanwhile we are aware of the capability of the framework used in the `bain` package - especially the opportunity to use multiply imputed data (Hoijtink, Gu, Mulder, & Rosseel, 2019). We therefore created another preregistration.

Data collection procedures

Our goal is to obtain a sample from the population of student teachers or teachers. This population is specifically suited to study the effect of Open Science Practices on trustworthiness, because it is part of their job to engage in evidence-based practice and thus stay up to date with research (Munthe & Rogne, 2015).

We plan to pass the data collection on to the Leibniz Institute for Psychology Information (ZPID).

Sample size Our design analysis implies that for a $d = .3$ and a BF of 3 or $\frac{1}{3}$ a sample of $N = 250$ is sufficient.

Sample size rationale

First preregistration:

Due to values missing by design, we approached data analysis and design analysis via two Bayes factor t-tests respectively. For design analysis we used the `BFDA` package. Required sample size from first preregistration: For small to medium effect, stopping rule of Bayes Factor of 10 ($\frac{1}{10}$ respectively) and 80% Power were N= 220. We thus aimed for a $N = 250$ with optional stopping at $BF = 10$ or $\frac{1}{10}$ respectively. Due to expected variations in the BF with low n, we proposed to begin data observation at $n = 150$.

Current Preregistration:

With the `bain` package we were able to tailor simulation to our design and specify informative hypotheses to be compared.

We conducted design analysis for Bayesian repeated measures analysis (one within factor) with missing data.

Based on the results of the first preregistration, we used $N = 250$ as sample size for power analyses with informed hypothesis approach, too. The further settings were $d = .3$ and $BF = 3$ or $\frac{1}{3}$.

```
####################################################################
# Bayesian repeated measures analysis (one within factor) with ###
# missing data                                                 ###
#                                                              ###
# A design analyses for the project re-buildging trust         ###
# assuming for small effect according Cohen (1988)             ###
####################################################################

library(bain)
library(psych)
library(MASS)
library(mice)
library(tidyverse)
library(hrbrthemes)
library(data.table)

sim_n <- 1000   # number of studies to simulate
true_d <- .3    # size of Cohen's d if mean_i != mean_j

# initialize data frame to store results in
sim_results_total <- tibble(
```

```
  true_hyp = character(),
  study_iteration = integer(),
  numerator = character(),
  denominator = character(),
  BF = numeric(),
  N = numeric())


#################################################################
## Loop over N
for(N in c(150)){

#################################################################
## Loop over study
for(study_iteration in 1:sim_n){

#################################################################
## Loop over true effects

for(true_eff in c("nonosp=nonvis=visosp",
                  "nonosp<nonvis<visosp",
                  "nonosp<nonvis=visosp",
                  "nonosp,nonvis,visosp")){


# Generate the data #################################
data <- data.frame(mvrnorm(n=N,
                           mu = if(true_eff == "nonosp=nonvis=visosp")
                                c(0,0,0) else
                                if(true_eff == "nonosp<nonvis<visosp")
                                c(-true_d,0,true_d) else
                                if(true_eff == "nonosp,nonvis,visosp")
                                c(true_d, 0, -true_d) else
                                c(-true_d,0,0),
                           Sigma = matrix(c( 1, .5, .3,
                                            .5,  1, .5,
                                            .3, .5,  1),
                                          3, 3)))
names(data) <- c("nonosp","nonvis","visosp")

# Generate missing values
data$nonosp[(0*floor(N/3)+1):(1*floor(N/3))] <- NA
data$nonvis[(1*floor(N/3)+1):(2*floor(N/3))] <- NA
data$visosp[(2*floor(N/3)+1):(3*floor(N/3))] <- NA

# Impute the data multiply ##########################
 M <- 100 # number of imputed data sets
 out <- mice(data = data, m = M,
             meth=c("norm","norm","norm"),
             diagnostics = FALSE,
```

```r
             printFlag = FALSE)

# Set up the matrices for the estimates ##############
mulest <- matrix(0,nrow=M,ncol=3) # setup of matrices
                                  # to store multiple estimates
covwithin <- matrix(0,nrow=3,ncol=3) # and covariance matrices


# Estimate the coefficients for each data frame ######
for(i in 1:M) {
 within <- lm(cbind(nonosp,nonvis,visosp)~1, # estimate the means
              data=mice::complete(out,i))    # of the three variables
 mulest[i,]<-coef(within)[1:3]           # store these means in
                                         # the matrix `mulres`
 covwithin<-covwithin + 1/M * vcov(within)[1:3,1:3]  # compute the
}                               # average of the covariance matrices


# Compute the average of the estimates ###############
estimates <- colMeans(mulest)
names(estimates) <- c("nonosp","nonvis","visosp")
covbetween <- cov(mulest) # is this the between covariance matrix?
covariance <- covwithin + (1+1/M)*covbetween # is this the
                                             # total variance?


# Determine the effective and real sample sizes ######
samp <- nrow(data) # real sample size
nucom<-samp-length(estimates)

# corresponds to Equation (X) in Hoijtink, Gu, Mulder, & Rosseel (2019)...
lam <- (1+1/M)*(1/length(estimates))*
  sum(diag(covbetween %*% ginv(covariance))) # ... (43)
nuold<-(M-1)/(lam^2) # ... (44)
nuobs<-(nucom+1)/(nucom+3)*nucom*(1-lam) # ... (46)
nu<- nuold*nuobs/(nuold+nuobs) # ... (47)
fracmis <- (nu+1)/(nu+3)*lam + 2/(nu+3) # ... (48)
neff<-samp-samp*fracmis
# coerce `covariance` to a list
covariance<-list(covariance)


# Test the hypotheses with bain ######################
results <- bain(estimates,
             "nonosp=nonvis=visosp;nonosp<nonvis<visosp;nonosp<nonvis=visosp",
             n=neff,Sigma=covariance,
             group_parameters=3,joint_parameters = 0)

sim_result <- tibble(true_hyp = true_eff,
                     study_iteration = study_iteration,
```

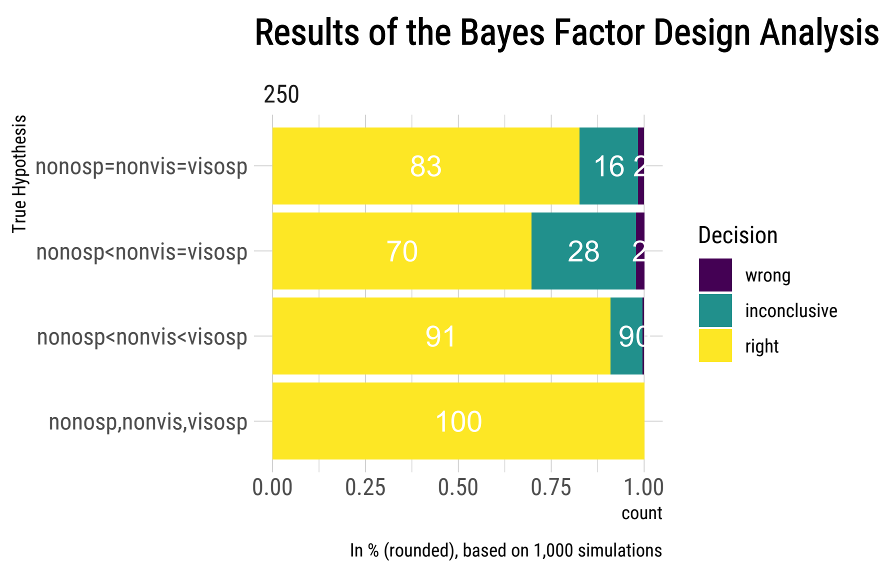
```
                      numerator = results$hypotheses,
                      denominator = "Hc",
                      BF = results$fit$BF[1:3],
                      N = nrow(data))%>%
  full_join(tibble(true_hyp = true_eff,
                    study_iteration = study_iteration,
                    numerator = results$hypotheses,
                    `nonosp=nonvis=visosp` = results$BFmatrix[, 1],
                    `nonosp<nonvis<visosp` = results$BFmatrix[, 2],
                    `nonosp<nonvis=visosp` = results$BFmatrix[, 3],
                    N = nrow(data))%>%
              gather(denominator, BF,
                      `nonosp=nonvis=visosp`,
                      `nonosp<nonvis<visosp`,
                      `nonosp<nonvis=visosp`))%>%
  filter(numerator != denominator)


 sim_results_total <- full_join(sim_results_total, sim_result)
}
  print(paste(N, study_iteration, sep = "_"))
 }
 }

write_csv(sim_results_total, "sim_results_total_bfda_badgestudy.csv")
```

```
## Recoding the results ############

#  1) Check the BF's for the comparisons of the true hypothesis (A)
#     against the other hypotheses under consideration and it's
#     complement.
#  2) If the data favors the true hypothesis against all others
#     under consideration and it's complement with a BF > 3
#     code Â»evidence for the true hypothesisÂ«
#  3) If this procedure results in at least one BF with
#     1/3 < BF < 3 code Â»inconclusiveÂ«
#  4) If this procedure results in at least one BF < 1/3 code
#     Â»wrong resultÂ«

library(tidyverse)
library(hrbrthemes)

## Recoding right and (inconclusive or wrong) decisions if
## `nonosp=nonvis=visosp` is true
`results_nonosp=nonvis=visosp_true` <-
  read_csv(("sim_results_total_bfda_badgestudy.csv"))%>%
  filter(true_hyp == "nonosp=nonvis=visosp")%>%
  group_by(N, study_iteration)%>%
  do(data.frame(decision = ifelse(
    filter(., numerator == "nonosp=nonvis=visosp" &
             denominator == "Hc")$BF > 3 &
```

```
    filter(., numerator == "nonosp=nonvis=visosp" &
             denominator == "nonosp<nonvis<visosp")$BF > 3 &
    filter(., numerator == "nonosp=nonvis=visosp" &
             denominator == "nonosp<nonvis=visosp")$BF > 3,
    "right",
    ifelse(
      filter(., numerator == "nonosp=nonvis=visosp" &
               denominator == "Hc")$BF > 1/3 &
      filter(., numerator == "nonosp=nonvis=visosp" &
               denominator == "nonosp<nonvis<visosp")$BF > 1/3 &
      filter(., numerator == "nonosp=nonvis=visosp" &
               denominator == "nonosp<nonvis=visosp")$BF > 1/3,
      "inconclusive", "wrong")),
    true_hyp = .$true_hyp[1]))

## Recoding right and (inconclusive or wrong) decisions if
## `nonosp<nonvis<visosp` is true
`results_nonosp<nonvis<visosp_true` <-
  read_csv(("sim_results_total_bfda_badgestudy.csv"))%>%
  filter(true_hyp == "nonosp<nonvis<visosp")%>%
  group_by(N, study_iteration)%>%
  do(data.frame(decision = ifelse(
    filter(., numerator == "nonosp<nonvis<visosp" &
             denominator == "Hc")$BF > 3 &
    filter(., numerator == "nonosp<nonvis<visosp" &
             denominator == "nonosp=nonvis=visosp")$BF > 3 &
    filter(., numerator == "nonosp<nonvis<visosp" &
             denominator == "nonosp<nonvis=visosp")$BF > 3,
    "right",
    ifelse(
      filter(., numerator == "nonosp<nonvis<visosp" &
               denominator == "Hc")$BF > 1/3 &
      filter(., numerator == "nonosp<nonvis<visosp" &
               denominator == "nonosp=nonvis=visosp")$BF > 1/3 &
      filter(., numerator == "nonosp<nonvis<visosp" &
               denominator == "nonosp<nonvis=visosp")$BF > 1/3,
      "inconclusive", "wrong")),
    true_hyp = .$true_hyp[1]))


## Recoding right and (inconclusive or wrong) decisions if
## `nonosp<nonvis=visosp` is true
`results_nonosp<nonvis=visosp_true` <-
  read_csv(("sim_results_total_bfda_badgestudy.csv"))%>%
  filter(true_hyp == "nonosp<nonvis=visosp")%>%
  group_by(N, study_iteration)%>%
  do(data.frame(decision = ifelse(
    filter(., numerator == "nonosp<nonvis=visosp" &
             denominator == "Hc")$BF > 3 &
    filter(., numerator == "nonosp<nonvis=visosp" &
```

```
              denominator == "nonosp=nonvis=visosp")$BF > 3 &
    filter(., numerator == "nonosp<nonvis=visosp" &
              denominator == "nonosp<nonvis<visosp")$BF > 3,
    "right",
    ifelse(
      filter(., numerator == "nonosp<nonvis=visosp" &
                denominator == "Hc")$BF > 1/3 &
      filter(., numerator == "nonosp<nonvis=visosp" &
                denominator == "nonosp=nonvis=visosp")$BF > 1/3 &
      filter(., numerator == "nonosp<nonvis=visosp" &
                denominator == "nonosp<nonvis<visosp")$BF > 1/3,
      "inconclusive", "wrong")),
    true_hyp = .$true_hyp[1]))


## Recoding right and (inconclusive or wrong) decisions if
## `nonosp,nonvis,visosp` is true
`results_nonosp,nonvis,visosp_true` <-
  read_csv(("sim_results_total_bfda_badgestudy.csv"))%>%
  filter(true_hyp == "nonosp,nonvis,visosp")%>%
  group_by(N, study_iteration)%>%
  do(data.frame(decision = ifelse(
    filter(., numerator == "nonosp=nonvis=visosp" &
              denominator == "Hc")$BF < 1/3 &
    filter(., numerator == "nonosp<nonvis=visosp" &
              denominator == "Hc")$BF < 1/3 &
    filter(., numerator == "nonosp<nonvis<visosp" &
              denominator == "Hc")$BF < 1/3,
    "right",
    ifelse(
      filter(., numerator == "nonosp=nonvis=visosp" &
                denominator == "Hc")$BF < 3 &
      filter(., numerator == "nonosp<nonvis=visosp" &
                denominator == "Hc")$BF < 3 &
      filter(., numerator == "nonosp<nonvis<visosp" &
                denominator == "Hc")$BF < 3,
      "inconclusive", "wrong")),
    true_hyp = .$true_hyp[1]))


## Joining the results_...._true tables
results_labeled <- full_join(
  `results_nonosp<nonvis<visosp_true`,
  full_join(`results_nonosp<nonvis=visosp_true`,
            full_join(`results_nonosp=nonvis=visosp_true`,
                      `results_nonosp,nonvis,visosp_true`)))%>%
  mutate(Decision = factor(decision,
                           levels = c("wrong",
                                      "inconclusive",
                                      "right")))
```

```
## Visializing the results
ggplot(results_labeled, aes(true_hyp, fill = Decision)) +
  geom_bar(position = "fill") +
  geom_text(aes(label=round(..count../1000*100), y= ..count../1000),
            position =position_stack(vjust = 0.5), stat= "count",
            color = "white", size = 5) +
  coord_flip() +
  facet_wrap(~N, ncol = 1) +
  labs(title = "Results of the Bayes Factor Design Analysis",
       caption = "In % (rounded), based on 1,000 simulations") +
  xlab("True Hypothesis") +
  scale_fill_viridis_d() +
  theme_ipsum_rc()
```



The results reveal good power under all hypotheses with the exception of $\mu_1 < \mu_2 = \mu_3$, where we find moderate power. Nevertheless, ...
The results also reveal low probability of false-positive results with $N = 250$, which justifies using a $BF = 3$ or $\frac{1}{3}$.

Stopping rule Based on the Bayes factor design analysis we aim at $N = 250$. A sample size slightly over $N = 250$ might still be possible: We didn't implement automated stopping, but have to manually check the sample size and stop the online survey.

Variables

Manipulated variables

Parallel to the first preregistration:

There are three conditions:

1. visOSP condition: Subjects receive a title page of an empirical study (Title, Abstract, Keywords, Introduction, ...) together with three Open Science badges. The badges are explained using hints in style of speech bubbles and indicate that the authors engaged in the OSP open data, open analysis script and open materials.
2. nonvis condition: Subjects receive a title page of an empirical study (Title, Abstract, Keywords, Introduction, ...) with no further information on Open Science, reflecting a "standard" journal article.
3. nonOSP condition: Subjects receive a title page of an empirical study (Title, Abstract, Keywords, Introduction, ...) together with three Open Science badges. The badges are explained using hints in style of speech bubbles and indicate that the authors did not engage in the OSP open data, open analysis script and open materials.

See survey here: https://osf.io/fh37z/

Measured variables

Parallel to the first preregistration:

- Trustworthiness: We apply the Muenster Epistemic Trustworthiness Inventory (Hendriks, Kienhues, & Bromme, 2015) with all three subscales. However as dependent variable we will only employ the subscale integrity. The other two subscales may be used for further exploratory analyses.

- Topic-specific multiplism: We apply an established scale on the topic specific multiplism (Merk, Rosman, Rueß, Syring, & Schneider, 2017).

- Topic-specific consistency: We apply the stablished three item-measure (Merk et al., 2017)

- Treatment check (treatment-specific): We mesure the perceived openness/ transparency of the empirical study via specifically developed items.

- Treatment check (global): We assess whether participants evaluate explanations of badges to be comprehensible, whether participants read the explanations and whether they perceive the explanations had an effect on their evaluations of authors.
- Additional small set of demographic variables will be assessed.

For detailed insights, see survey here: https://osf.io/fh37z/

Indices We are going to built sum scores for the METI dimensions. Furthermore we will exploratory investigate the measurement invariance of the METI.

Analysis Plan

Statistical models Analyses will be conducted parallel to design analyses (see script above). If the data fails to meet assumptions, we plan to apply robust alternative analyses (Bosman, 2018).

Transformations None planned.

Inference criteria We will use Bayes factors with thresholds of 3 (or $\frac{1}{3}$ respectively), based on the design analysis.

Data exclusion Implausible (consistent), out of theoretical range responses and participants taking less than 5 minutes for the survey may be eliminated for the analyses. The reasoning and decision to eliminate these participants will be made prior to data analysis and reported in disseminations.

Missing data	*Parallel to the first preregistration:* Multiple imputation will be used.
Exploratory analyses (optional)	*Parallel to the first preregistration:* • Hypothesis 3: BF Moderation Analysis will be conducted with visible OSP (vs. not visible vs. visibly non-OSP) as predictor, topic specific multiplism as moderator and perceived trust (subscale integrity) as dependent variable • Hypothesis 4: BF analysis with visible OSP (vs. not visible vs. visibly non-OSP) as predictor and topic specific multiplism as dependent variable will be computed

Other

Other (Optional)	

References

Bosman, M. (2018). *Robust Bayes factors for Bayesian ANOVA: Overcoming adverse effects of non-normality* (PhD thesis). Utrecht University, Utrecht.

Geukes, K., Schönbrodt, F. D., Utesch, T., Geukes, S., & Back, M. D. (2016). Wege aus der Vertrauenskrise: Individuelle Schritte hin zu verlässlicher und offener Forschung. *Zeitschrift Für Sportpsychologie*, *23*(3), 99–109. doi:10.1026/1612-5010/a000167

Gu, X., Hoijtink, H., Mulder, J., & Lissa, C. J. van. (2019). *Bain: Bayes Factors for Informative Hypotheses.*

Hendriks, F., Kienhues, D., & Bromme, R. (2015). Measuring Laypeople's Trust in Experts in a Digital Age: The Muenster Epistemic Trustworthiness Inventory (METI). *PLOS ONE*, *10*(10). doi:10.1371/journal.pone.0139309

Hoijtink, H., Gu, X., Mulder, J., & Rosseel, Y. (2019). Computing Bayes factors from data with missing values. *Psychological Methods*, *24*(2), 253–268. doi:10.1037/met0000187

Lupia, A. (2018). The Role of Transparency in Maintaining the Legitimacy and Credibility of Survey Research.

In D. L. Vannette & J. A. Krosnick (Eds.), *The Palgrave Handbook of Survey Research* (pp. 315–318). Cham: Springer International Publishing. doi:10.1007/978-3-319-54395-6_41

Merk, S., & Rosman, T. (2018). *Smart but evil? Student-teachers perception of educational researchers' epistemic trustworthiness.* (Preprint). Open Science Framework. doi:10.31219/osf.io/cduqe

Merk, S., Rosman, T., Rueß, J., Syring, M., & Schneider, J. (2017). Pre-service teachers' perceived value of general pedagogical knowledge for practice: Relations with epistemic beliefs and source beliefs. *PloS One*, *12*(9), 0184971. doi:10.1371/journal.pone.0184971

Munthe, E., & Rogne, M. (2015). Research based teacher education. *Teaching and Teacher Education*, *46*, 17–24. doi:10.1016/j.tate.2014.10.006

Wingen, T., Berkessel, J., & Englich, B. (2019). *No Replication, no Trust? How Low Replicability Influences Trust in Psychology* (Preprint). Open Science Framework. doi:10.31219/osf.io/4ukq5