

Avila CREP Replication of Cameron et al. (2019)

Sydney Gardner & Jordan Wagge

2023-12-07

Contents

Introduction	1
Background	2
Replication	2
The Replication Crisis	2
Collaborative Replication	2
Cameron et al.'s Study	3
The Present Study	4
Extension Hypotheses	4
Method	4
Participants	4
Procedures and Materials	4
Results	5
NASA Scores	5
rewrite below – we now <i>do</i> find a significant correlation between deck choice and perceived effort	6
Time Course Interactions	6
we now do have a significant main effect of time	6
Extension Hypotheses	8
this last test is no longer significant	8
Discussion	8
Limitations	8
Future Research	9
References	9

Introduction

Replicating a study gets a bad rap in science as many see this process as repeating a study's procedures exactly, ignoring the benefits of this process. Replication is important to science as it establishes credibility for their findings with new data as evidence (Nosek & Errington, 2020). Not only can researchers see if the findings reoccur, following the same methodology, but they can also explore additional areas. Researchers believe that replication is just being redundant or stating the obvious if you happen to have the same findings, but in reality, these findings are most significant because they strengthen the credibility of existing findings. Additionally, replication of previous findings improves generalizability, even under conditions that will differ from the original study (Nosek & Errington, 2020).

Background

Replication

Replication is often utilized within research, reproducing or repeating a previous study's methods and procedures. This approach is used to see if the previous findings may be generalizable. This helps to prove reliability and validity of previous research. It's important to practice replication in order to see if the original findings were a one-time result, or if in-fact this phenomena is consistent and generalizable. Replication does not always result in similar findings due to different conditions, such as different samples, settings, etc. Of course, a sample is never able to be exactly replicated, so findings consistent among other samples within these replications again helps to see if these findings are reliable and valid. Lastly, it is not a bad thing for a replicated study to not result in similar findings. Replication also allows for researchers to discover new or additional findings. Additionally, things that are wrong in the previous research may be found. Replication is an opportunity to identify any errors or limitations in the previous research.

Replication plays a significant role in research. Although novel findings are significant to address, the ability to replicate another researcher's findings only helps the credibility of said findings. Since so many factors could have made the findings a one-time event, it's important to see if these findings could be replicated in another environment. Additionally, it's important to see if further significant findings could be explored through additional measures added or if better methodology could be utilized. Despite the benefits of replication, oftentimes researchers don't want to be involved in replication studies since it is highly tempting to be involved in novel research. Also, people are more inclined to participate in novel research since they want to discover new findings, or perhaps don't want to go through all the work of the experiment to get expected results.

The Replication Crisis

An ongoing issue within the replication of studies is the "replication crisis". The replication crisis remains a prominent issue as much of psychology and its findings may not be reliable or accurate. When attempts are made to replicate previous findings, these studies may fail to reproduce the results of the previous research. This can lead to concern or doubt about the validity of the previous studies. The promotion of replication remains important to addressing the replication crisis by contributing to the reliability of previous findings. The replication crisis is rooted by many problems including data fabrication, data misrepresentation, publication bias or self-censoring, and post hoc hypothesizing. Data fabrication in research can occur when researchers alter their data to support their predictions, making these outcomes more difficult to replicate. Researchers could have also misrepresented their data by cherry-picking results or omitting contradictory data.

Additionally, publication bias may affect replication since researchers may show only positive or significant results, disregarding nonsignificant or negative findings that may come up in the replication as well. Lastly, many researchers fall victim to post hoc hypothesizing, or modifying their hypotheses after analyzing the data, which makes it difficult for previous findings to be replicated. Any of these practices affects the reliability of research and can impede future replications.

Collaborative Replication

CREP, otherwise known as the Collaborative Replication and Education Project, allows for undergraduate students and the corresponding faculty to participate in direct replications of selected studies (Wagge et al., 2019). These replications prove to be educational to students as they are taught good scientific practices through performing direct replications on high quality cited articles, while applying their research skills when using open science methods. CREP proves to be an excellent tool for promoting replication since data from direct replications help establish credibility for the discipline. It also benefits students as they are receiving training in beneficial research practices. CREP is inclusive to those with limited knowledge or expertise, therefore, students aren't required to have extensive knowledge of a literature or research design to participate in these replications; instead they gain this knowledge along the way.

CREP studies are selected for crowdsourced direct replications by the CREP team. They are chosen from

samples of the most highly cited papers from majorly cited journals. These articles are curated to be presented to faculty or research teams that will facilitate these replications with the help of their students or fellow researchers. It is also important that these research processes are plausible to be replicated in an institutional or lab setting that students can partake in the replication. To complete CREP projects, the student is first introduced to CREP by a faculty instructor at their institution. They work together to coordinate specific tasks or assignments and meet deadlines during the replication process. Although student groups typically conduct direct replications of the selected studies, it is possible to include additional measures or conditions that may be investigated as they test their own exploratory hypotheses. This replication process allows students to gain skills in following research methods and utilizing critical evaluation skills to investigate scientific claims.

When students become involved in a CREP project, faculty often instruct them on their assigned duties and guide them through each step of the replication process. Students can first get involved by reading the empirical paper of the selected study. After this, students and their faculty mentors can decide if their methods will be a direct replication of the study, or if they wish to include any extension hypotheses. Students can also get practice at filling out an IRB application that will be submitted and reviewed to the CREP team. Once approved, students will then be able to collect and analyze data, later submitting the results for final review by the CREP team. Students may choose to complete a final project such as a poster or presentation to share their findings, even if not consistent with the original study.

CREP can lead students to publication of their work, compared to other teaching models that are much lengthier to approve every step of a student's experimental design or creating their own hypotheses or collecting data, with no guidance towards sharing their research. CREP helps students get first-hand experience with scientific dissemination since students become the primary researchers that can not only participate in the replication and collecting data, but also in presenting or sharing their results in presentations or even at conferences. Open Science Framework (OSF) is also a highly beneficial tool for students to practice sharing their data to other researchers. Students can also learn to not only collect data and test hypotheses, but also how to write up and present this data, participating in the authorship process. These replications can be more beneficial to students since it may be more likely to result in a publication for the student than other research models. Even if the direct replications come up with unexpected results, it is still worth sharing and viable for a student to be published. The process can also be completed faster than other research models since they are designed and selected to be completed with a typical semester. CREP is designed to be a more efficient and productive method of promoting and applying research practices for students and other researchers, especially those who could gain research skills from it.

Cameron et al.'s Study

In Cameron et al.'s (2019) study on empathy, they found that participants were more aversive to, or avoidant of empathy when they found it to be effortful, or if perceived cognitive costs of empathy were present. This research is important to explore further as empathy is a good thing for individuals to have, but often they fail to utilize it, or have it in many situations. Previous research only focused on material and emotional costs of empathy, and this study chose to expand on this with the examination of how cognitively tasking individuals view empathy and how this may affect their desire to empathize. To explore this, they had participants recruited on MTurk complete the Empathy Selection Task, having participants choose from binary choices and selecting situations after viewing a photo of a child refugee. Participants viewed different decks with various instructions on their experiences with empathy. Participants were tasked with choosing between a "DESCRIBE" and a "FEEL" deck displayed on their screen. The depiction of the child refugee is then shown with the corresponding prompt for their selected deck. Participants completed 40 trials to assess what effect perceived cognitive costs may have on their desire to empathize, responding accordingly to their selected deck each time. This study found that individuals perceived the empathy related deck as more effortful and aversive, demonstrating that empathy efficacy increased a participant's likelihood of empathizing in the situation. Participants displayed more empathy avoidance when there was perceived cognitive costs. An individual's empathy increases when they interpret the situation as less cognitively costly.

The Present Study

Following the procedures of Cameron et al.'s study, we hypothesize that the present study will replicate their previous findings. We expect that participants will demonstrate a preference for the empathy deck when completing the trials of the Empathy Selection Task.

Extension Hypotheses

As an extension hypothesis, our class decided to test whether self-reported fatigue and deck choice were associated. Specifically, we believed that participants who chose empathy more would express higher levels of mental – but not physical – fatigue after the study was complete.

Method

Participants

For this replication, participants were recruited from an undergraduate research pool at Avila University, using convenience sampling. Additionally, this study was advertised on social media and shared by researchers on this project. Our initial N was 159. The present study had a total of 79 participants, with 80 participants excluded from consideration when analyzing our data because of incomplete or test data. Criteria for exclusion from this study was determined based on if the individual failed to complete the entire survey, and 79 participants were removed for that reason while one test case participant was removed.

In order to participate in this study, participants had to be at least 18 years of age ($M = 2462.14$, $SD = 21598.53$), with participants ranging between ages 18 and 75. The majority of participants (72.15%) identified as female, in addition to 17.72% identifying as male, and 1.27% identifying as non-binary. Demographics were also collected for participants' race/ethnicity, finding that the majority of participants (60.76%) identified as white. Additionally, 7.59% of participants were Black, 11.39% were Hispanic/Latino, 1.27% were Indian, 1.27% were Indigenous, and 2.53% were multiracial (responded with multiple racial/ethnic identities).

Procedures and Materials

Participants completed this online study on SoSci Survey. Participants were told beforehand that they should be able to complete the survey within 30 minutes. After completing the informed consent form, participants were instructed that they would complete a series of trials, seeing two decks of cards: the red “DESCRIBE” deck on the left, and the blue “FEEL” deck on the right. Participants were asked to choose between these decks. Once a deck has been chosen, an image of a person is revealed. Depending on the deck chosen by the participant, they were given one of two possible sets of instructions.

When a participant selects from the “DESCRIBE” deck, the participant was told to look at the image objectively, focusing on the external features and appearances of the person in the image. For these trials, participants were asked to type one sentence describing the age and gender of the person in the image. When a participant selects from the “FEEL” deck, the participant was told to look at the image empathetically, focusing on the internal feelings and experiences of the person in the image. For these trials, participants were asked to type one sentence describing the feelings and experiences of the person in the image. Participants were able to independently choose from either deck on any trial, and were free to choose from each deck at any time. Participants were not required to choose any deck for a specific amount of trials, and could in fact choose to favor a deck, completing more trials on it than the other deck.

After completing the series of trials, participants were tasked to complete a series of questions with open-ended responses, assessing their experience with the Empathy Selection Task. The questions asked participants to share how they chose between decks, or if they had any preferences, in addition to if they noticed any differences between the decks. Participants were then asked to answer questions regarding both the “DESCRIBE” and “FEEL” deck. In regard to both decks, participants answered questions using a 5-point Likert scale, ranging from “Very Low” to “Very High”. Some questions included “How mentally demanding was this deck?”, and “How successful were you in accomplishing what you were asked to do in this deck?”.

Following this, participants were asked to respond to demographic questions about their gender identity, racial/ethnic identity, and age. Additionally, participants were asked to complete a few multiple-choice questions about their political orientation (ranging from extremely liberal to extremely conservative, as well as how religious they are (ranging from not at all religious to extremely religious)).

Next, participants looked at an image of a stack of coins and were told to think of it as representing where people stand in the United States. They were then asked, “Where would you place yourself on this stack of coins?” by moving the “X” on the screen to where they think they stand. Finally, to gauge participant fatigue, they completed the Multidimensional Fatigue Inventory (Smets et al., 1995), a 20-item survey that assesses self-reported fatigue for five different categories: mental, physical, general, reduced motivation, and reduced activity. Participants answered a series of questions using a 7-point Likert-type scale, ranging from “yes, that is true” to “no, that is not true”.

Results collected on SoSci were downloaded, evaluated, then analyzed once the desired number of participants had completed the study. Participant responses were reviewed to ensure that participants responded appropriately to the chosen deck. Responses were coded by graduate student researchers involved in this replication study, under the supervision of their faculty mentor. All responses were reviewed and flagged if a participant didn’t respond appropriately to their selected deck (e.g., failing to describe how the child looked for the “DESCRIBE” deck; failing to describe how the child felt for the “FEEL” deck). After reviewing all the data, 4.32% of participant responses for the “DESCRIBE” or “FEEL” trials were flagged by the coders, but remained included in the data analyzed.

Results

Upon completion of reviewing all participant responses, in order to assess participant preference between the two decks, we employed the following statistical analyses. We conducted a one-sample t-test against a value of .5, in order to determine whether participants tended to prefer one deck over the other (empathy = 1, neutral/describe = 0). In the original study, they found significant differences in the mean empathy choice when compared to .5, $t(55) = -4.70$, $p < .001$. These results indicated that more participants favored the neutral deck ($M = .33$, $SD = .27$). In the present study, we were unable to replicate this effect, with our findings showing no significant differences, $t(78) = -1.07$, $p = 0.286$. Our analysis indicated that participants did not exhibit a preference for one deck over the other ($M = 0.47$, $SD = 0.24$).

In the original study, after conducting a meta-analysis over 11 different studies they found that the empathy deck, when compared to the objective, was characterized as more effortful, less efficacious, and more aversive. These differences across all studies examined were statistically significant ($p < .001$).

In the present study, we conducted similar one-sample t-tests using our data collected.

NASA Scores

Effort To calculate the effort score, we subtracted effort scores for the blue deck (“feel”) from effort scores for the red deck (“describe”). Numbers above 0 indicate greater effort for the red (describe/objective) deck, while numbers below 0 indicate greater effort for the blue (empathy) deck. Scores for effort were significantly below 0 ($M = -0.85$, $SD = 1.36$), indicating that participants perceived the empathy deck to require more effort ($t(78) = -5.58$, $p < .001$, $d = -0.63$).

Aversiveness We computed aversiveness scores in the same manner (subtracting blue scores from red scores, so that numbers lower than 0 indicate more aversiveness for the blue empathy deck). Participants did rate the blue deck as more aversive than the red deck ($M = -0.8$, $SD = 1.43$), $t(78) = -4.97$, $p < .001$, $d = -0.56$.

Efficacy Similarly, we computed efficacy scores and found that participants felt the red (objective) deck was more efficacious than the blue (empathy) deck ($M = 0.39$, $SD = 1.27$), $t(78) = 2.76$, $p = 0.007$, $d = 0.31$.

Deck selection and NASA subscales

rewrite below – we now *do* find a significant correlation between deck choice and perceived effort

Additionally, the original study also conducted a meta-analysis and found negative correlations between the selection of the empathy deck and ratings of effort and aversiveness associated with that deck. Again, the differences were all statistically significant ($p < .001$). In the present study, we conducted a Pearson's r correlation analysis to investigate if similar associations exist in our dataset. We found no significant correlation between deck choice (empathy selection) and perceived effort for each deck ($r = 0.23$, $p = 0.041$) or perceived aversiveness ($r = 0.11$, $p = 0.357$). However, a statistically significant negative correlation was found between deck choice and efficacy ($r = -0.45$, $p < .001$), indicating that participants who tended to choose the empathy deck more often than other participants also tended to perceive that deck as more efficacious (see Figure 1).

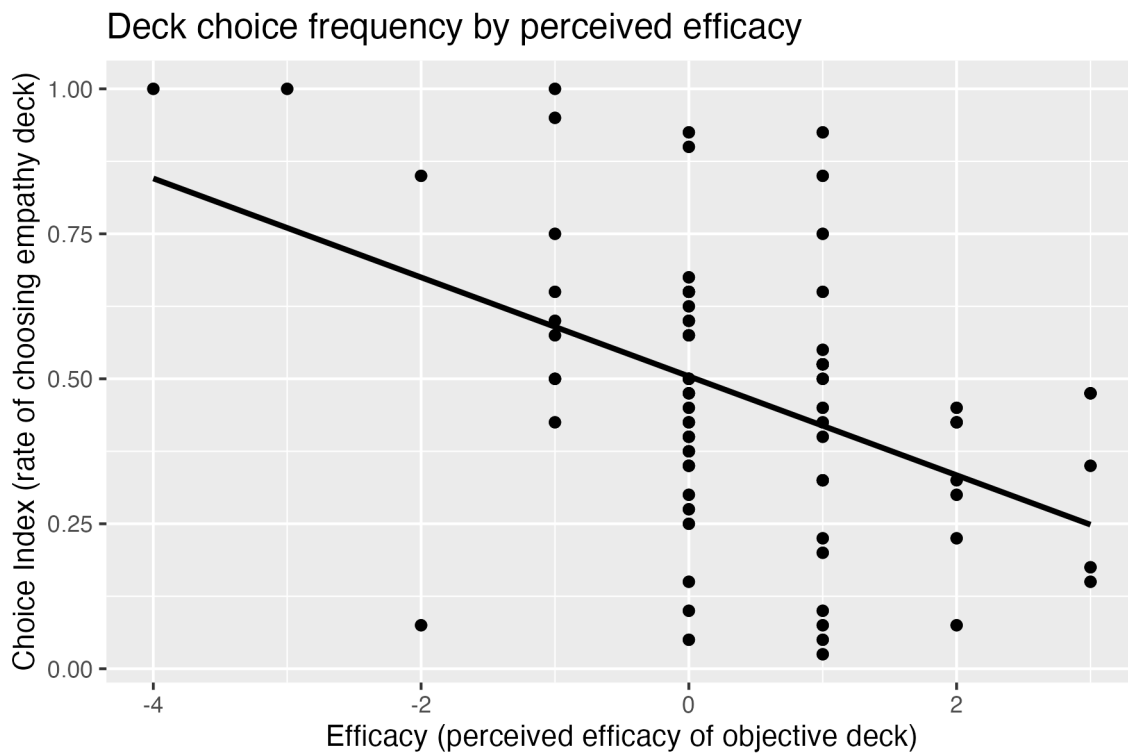


Figure 1: *Figure 1: Deck choice frequency by perceived efficacy.*

Time Course Interactions

we now do have a significant main effect of time

In the original paper, the authors examined whether there was an interactive effect on deck selection between time course (early versus late trials) and aversiveness, efficacy, and effort. We found no significant main effects of time ($\beta = -0.02$, $p = 0.03$) or interactions between time and effort ($\beta = 0.01$, $p = 0.1$) or time and aversiveness ($\beta = 0$, $p = 0.88$). However, we did find a significant interaction between time and efficacy, such that empathy deck selection appeared to decrease over time for participants who found the empathy deck less efficacious, and increase over time for participants who found the empathy deck more efficacious ($\beta = -0.02$, $p < .001$). This pattern is demonstrated in Figure 2:

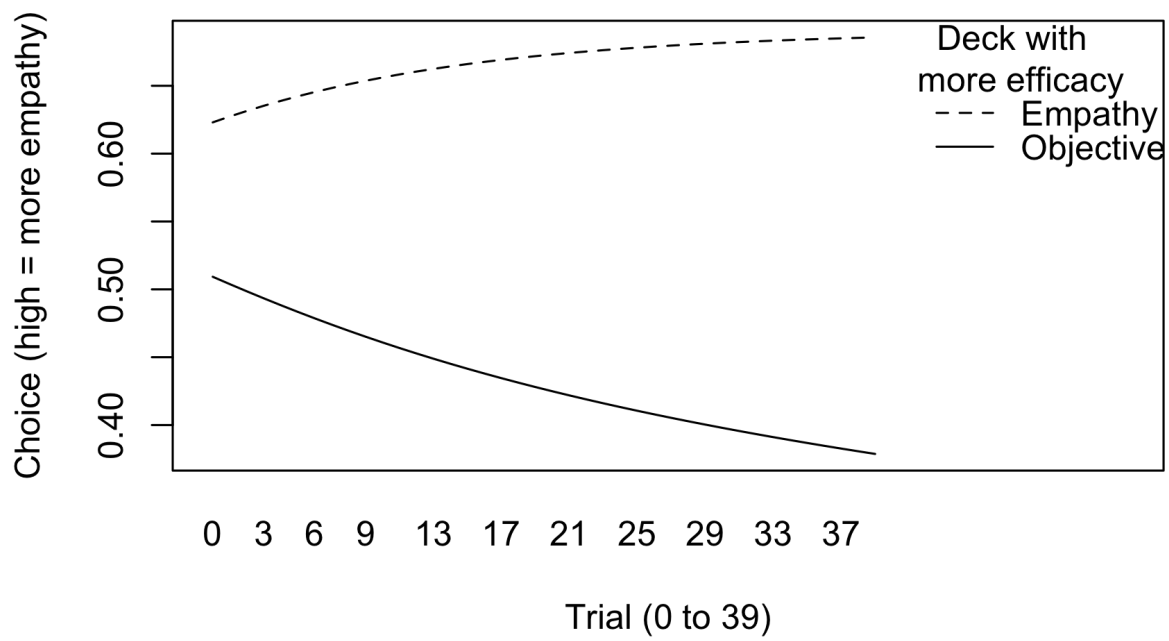


Figure 2: *Figure 2: Interaction between time (trial number) and efficacy.*

Extension Hypotheses

None of the fatigue subscales were significantly correlated with any of the indices. We specifically hypothesized that the number of times participants selected the empathy deck would be positively correlated with their mental, but not physical fatigue ratings. However, neither fatigue subscale was correlated with empathy deck selection (in order, r values: .14, -.13 and p values: .23, .27). In exploratory analysis, we found that there were no other significant correlations between the MFI subscales and deck choice with the exception of the subscale for reduced motivation; specifically, reduced motivation was significantly positively correlated with empathy deck selection, $r(76) = .10$, $t(76) = 0.9$, $p = .10$.

this last test is no longer significant

Discussion

Although we were unable to replicate the original study's findings, additional associations were found when examining participant preference when choosing between decks. In the present study, participants were not inclined to select one deck over the other, but characteristics such as perceived effort, efficacy, and aversiveness to the task may play a significant role in their decision making.

Previously found in Cameron et al.'s (2019) study, participants favored the neutral deck over the empathy choice. However, based on the results of the present study, when completing the Empathy Selection Task, participants did not demonstrate a preference for either deck ("DESCRIBE" or "FEEL") over the other. This difference could be a sign of individuals not being aversive to the cognitive costs of empathy, but it also may be a result of individual differences. One thing that may combat this theory is that consistently across both the original study and the present study, participants found the empathy deck to be more effortful, less efficacious, and more aversive.

Although the original study found that the more participants selected the empathy deck, the less effortful and aversive they perceived it to be, the present study did not find this same association. Our findings did not demonstrate a strong relationship between selection of the empathy deck and ratings of effort or aversiveness. However, our findings did demonstrate a strong relationship between selection of the empathy deck and greater perceived efficacy, meaning if participants thought that the empathy deck yielded better outcomes, they were more inclined to choose the empathy deck over the neutral deck.

While our extension hypotheses were not supported, exploratory analyses with the MFI found a positive correlation between reduced motivation and empathy deck selection. In other words, participants' indications of reduced motivation were linked to more empathy deck selection. This might suggest one of two or more different causal pathways, as it could be the case that frequent empathy deck selection reduces motivation *or* reduced motivation causes empathy deck selection. While the first explanation seems more likely, more research would be needed involved both pre- and post- measures in order to parse this relationship for a causal pathway.

Limitations

Since this study was completed online, there is a lack of control over the participant's environment or the setting in which they completed this survey. We have no way to ascertain if individual's avoided distractions that may have affected their attention throughout the series of trials they were tasked with completing. Additionally, since participants were recruited through an undergraduate research pool, or made aware of this study through advertisements locally or via social media, the employment of convenience sampling may hinder the sample from being representative of the general population. We also did not control for participant characteristics such as demographics, including socioeconomic status, educational background, etc. that may affect their deck preferences or ratings throughout this study. Lastly, self-reported measures were collected for the ratings of effort, efficacy, or aversiveness that they perceive when selecting the empathy deck. These types of responses may create a bias within participant responses as they might have responded to appease what they perceive the researchers are looking for, or perhaps responding as they think they should, rather than

responding with their true feelings. This may affect the reliability or validity of the construct, as participant responses may not be representative of their ratings.

Future Research

Many factors play a role in participants selection of a deck throughout these trials, therefore, examining individual differences in future replications, or a completely modified study, could be a novel approach at investigating how various factors may impact deck preference, in addition to perceived effort, efficacy, or aversiveness to the empathy deck. Exploration of these differences may explain inconsistent findings across replications, and may help to improve the generalizability of this research. Additionally, future research may make modifications to the decks to manipulate different aspects that may influence participants' deck preferences. These changes may be employed to see if variation in content, emotional intensity of the refugee images, etc. have any effect on their selection of the empathy deck. Finally, a longitudinal study design may better assess if participants' deck preference and ratings of effort, efficacy, and aversiveness are consistent over time. This could also be employed if a future replication did not replicate the original study's findings. Perhaps if observed at multiple instances over a certain time span, these insignificant findings could at least be displayed consistently, or significant findings may be observed at other instances within the longitudinal study.

References

- Cameron, C. D., Hutcherson, C. A., Ferguson, A. M., Scheffer, J. A., Hadjiandreou, E., & Inzlicht, M. (2019). Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *Journal of Experimental Psychology: General*, 148(6), 962–976. <https://doi.org/10.1037/xge0000595>
- Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS Biology*, 18(3), e3000691.
- Smets, E. M., Garssen, B., Bonke, B., & De Haes, J. C. (1995). The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research*, 39(3), 315–325. [https://doi.org/10.1016/0022-3999\(94\)00125-o](https://doi.org/10.1016/0022-3999(94)00125-o)
- Wagge, J. R., Brandt, M. J., Lazarevic, L. B., Legate, N., Christopherson, C., Wiggins, B., & Grahe, J. E. (2019). Publishing research with undergraduate students via replication work: The Collaborative Replications and Education Project. *Frontiers Psychology*, 10(247). <https://doi.org/10.3389/fpsyg.2019.00247>