**Peer review history**

**MS Title:** Reexamining the relationship between shift work and health behavior: Do fluid
intelligence, socio-economic status, and self-control moderate the relation?

**Author names:** Myriam A. Baum, Frank M. Spinath, and Elisabeth Hahn

**Editor first decision**

May 27, 2020

Dear Mrs Myriam Annabelle Baum,

Thank you for submitting your work to Collabra: Psychology. Foremost, I apologize for the delay in processing this paper. This was a busy semester and the pandemic hit. I am sorry for how long this paper has spent in review. Please accept my apology.

I was able to secure reviews from two outstanding scholars who are experts in personality and organizational psychology. These individuals are extraordinarily well qualified to review this paper and I thank them for their service to this journal. I independently read the paper and then consulted the comments.

As you will read below, the reviewers were positive about the topic and the sample. On the other hand, they had several concerns about the execution of the analyses and the interpretations. They have comments and suggestions that will improve the paper to a point where it is has a strong chance of publication at this outlet. Thus, I will extend a revise and resubmit decision. I do not plan to send the revision back to reviewers so the time to make a final decision should be reduced. However, I acknowledge that the suggested revisions are likely to be substantial.

I think both reviewers raised excellent points and I suspect both of their views are representative of the reactions of other readers to this paper. They did an outstanding job in my opinion and you should address each of their concerns either in the revised text or in the response letter. Addressing their concerns will make this a stronger paper.

To be honest, almost all of my concerns were echoed by these reviewers so there was strong agreement across all three readers of this paper. This made my job much easier. I will comment on the most salient points and offer some of my own reactions from reading the paper in my letter. You might disagree with some (or all) of the points so feel free to pushback against any suggestions you believe will harm your work. Please just describe your counterpoints in the return letter.

1. I think my first reaction when getting to the analyses was a sense of puzzlement as to why the sample was split by gender for the analyses. Both Reviewers also commented on this issue. We do not see a strong reason to split the sample and conduct separate analyses. I would simply include gender and gender by shift work interactions (as well as gender by other moderators) for all analyses.

2. Reviewer A provides several helpful references for thinking about effect sizes. I think placing more interpretation on the effect sizes and struggling with the issue of what would count as a practically significant or non-negligible effect size in both the Introduction and Discussion will enhance the paper.

3. I agree with Reviewer A who questioned the winsorizing approach. I also objected to the idea of “correcting” variables for age. Why not just include age in the models and even test for age by shift-work interactions? The term correcting seems to imply something that might not be intended. I suspect this is just convention from how some discuss twin-modeling decisions. Nonetheless, I think a different word other than “correct” would be preferable and a more direct approach to accounting for age would be even better.

4. I did not think the references to ego depletion were needed given the uncertain empirical status of that approach and the uncertain status of how that theory related to the specific “trait” measure of self-control used in this paper.

5. The measure of self-control was short. This might deserve comment and any information about convergent validity with other measures would be useful. What was the correlation across Wave 1 and Wave 2 and how many people had scores at both waves?

6. The idea of including a large correlation table suggested by Reviewer A strongly resonated with me.

7. Reviewer B had concerns about considering SES as an intelligence variable (see e.g., p. 26 in the paper). I strongly agree, and I think that is probably an unwise framing. I think it is fine to frame this as testing the moderating effects of SES, fluid intelligence, and self-control. Indeed, I would be quite specific about what moderators were tested (e.g., there was only one personality characteristic considered so it would be clearest to just talk about self-control rather than personality per se).

8. I like the constraints on generality statement paper from Simons, Shoda, and Lindsay (2017) and encourage you to include a dedicated section about such constraints in the Discussion. For example, I think non-German readers might benefit from an even broader perspective on German public policies about work and health care. In other words, the null results here might not apply to countries with different government policies and health care systems and I think additional discussion about boundary conditions is worthwhile. (Threads of such constraints are in the Discussion, so this would involve bringing them together and adding even more).

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. Perspectives on Psychological Science, 12(6), 1123-1128.

Those were the big issues based on my reading. I also noted many of the same issues as the Reviewers that I did not cover in #1 to #8 so please attend to all of their points if you revise this work. And I do hope you revise this paper given that I think it is interesting and worth getting into the literature.

To access your submission account, follow the below instructions:
1) login to the journal webpage with username and password
2) click on the submission title
3) click 'Review' menu option
4) download Reviewed file and make revisions based on review feedback
5) upload the edited file
6) Click the 'notify editor' icon and email the confirmation of re-submission and any relevant comments to the journal.

If you have any questions or difficulties during the revision process, please contact us. Thanks again for considering Collabra as a showcase for your interesting work. Feel free to contact me at donnel59@msu.edu if you have questions or concerns.

Sincerely,

Brent Donnellan
Michigan State University
donnel59@msu.edu

------------------------------------------------------
Reviewer A:

1) General comments and summary of recommendation
Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:
 - Are the methodologies used appropriate?
 - Are any methodological weaknesses addressed?
 - Is all statistical analysis sound?
 - Does the conclusion (if present) reflect the argument, is it supported by data/facts?
 - Is the article logically structured, succinct, and does the argument flow coherently?
 - Are the references adequate and appropriate?:
This paper presents an analysis of the relationship of shift work, cognitive ability, and measured self-control with self-reported health behaviors across various demographic groups. The paper uses a large sample. There are many strengths to the study, but as presented, there are several important issues with the data analysis and interpretation that should be addressed before publication.

Control variables and data handling
1. (p. 14) In your section reporting results of the MCAR test, please refer to the results as suggesting that male missing data “occurred \_completely\_ at random” and that female results “did not occur \_completely\_ at random”. Specifying „completely“ is important, as „Missing at Random“ has a different meaning (i.e., that some of the other predictors in the data can account for the missingness; this is the assumption you make with the female sample). However, Little’s MCAR test should generally not be used to determine a missing data handling strategy. This approach violates assumptions about the sampling distributions of the imputed data. If predictor variable are available that can potentially account for missingness, it is better to use these variables in multiple imputation, rather than to perform this MCAR test screening procedure.
2. You state that “Data were winsorized (Sheskin, 2003) – except for likert-scaled data (namely, health behavior, subjective health, and self-control) – and corrected for linear effects of age.”, but you do not provide any justification for these procedures. I see no reason to winsorize your data. Please report results using the observed data. Please provide a clear justification for controlling for age (cf. Spector & Brannick, 2011, https://doi.org/10.1177/1094428110369842). Please also report results not controlling for age.
3. Your regression tables state “For the female subsample, all variables were z-standardized before analyses.” Why only for the female sample? Please be consistent and report both unstandardized and standardized results for both samples. See below for comments on reporting and interpreting unstandardized results in this study because of the inherent meaningfulness of the measured outcome variables.

Interpretation of results
1. The paper focuses almost entirely on statistical significance in interpreting the results. This is a problem because most if not all of the reported effects appear to be practically negligible, even if they are „statistically significant“. Beyond whether the effects are statistically significant, please focus your discussion on the size of the effects. For example, I would look at all of zero-order shift work–outcome relationships reported on page 14 and conclude that shift work has negligible to small relationships with all health behvaiors, even the one that is nominally significant.

The work of (Funder & Ozer, 2019, https://doi.org/10.1177/2515245919847202) can be helpful in interpreting effects sizes. They reviewed distributions of effect sizes observed in psychological research and found that the 25th, 50th, and 75th percentiles for effect sizes correspond to correlation values of r = .10, .20, and .30, respectively. These might be interpreted as rough generic benchmarks for „small“, „medium“, and „large“ effects.

Additionally, all three of your criteria are meaningful in themselves. I recommend you focus on the \_unstandardized\_ regression coefficients as indices of the effects of shift work in terms of number of cigarettes smoked, amount of alcohol consumed, or estimated number of preventative heath behaviors. This „real-world“ anchoring of your results might better illustrate the meaningfulness (or not) of the impact of your predictors on these outcomes.

For the coefficients with fluid intelligence and self-control, you could partially standardize the coefficients only for the intelligence/self-control variable to estimat the effect of a 1 SD/15 IQ point difference.
2. Throughout the paper, you report statistical significance tests separately by group (e.g., separately for men and women for all analyses; differences in correlations between subjective health and behaviors across gender and shift work samples). As written, based on these analyses, you appear to conclude in various places that variables are related in one group, but not the others (e.g., you conclude that shift work is related to alcohol use in men, but not women; you conclude that subjective health is related to smoking in male day workers, but not other groups). These analyses do not permit this inference. A difference in statistical significance p value across groups \_does not imply\_ a statistically or practically significant difference in the relationship across groups. To make an inference that variable relationships differ across groups, you must (1) focus on the difference in estimated effect sizes across group and (2) formally test the differences in relationships.

For example, in the zero-order shift work–outcome relationships reported on p. 14–15, gender does not appear to be a substantial moderator of any result. However, as written, the discussion appears to suggest that results are different for men/women for the shift work–alochol relationship, which isn’t really the case. The effect appears to be fairly negligible in both cases. A rough calculation based on your reported β values and the group sample sizes shows me that the ∆β = .09 [95% CI .03, .15]. While this is technically „statistically significant“, the difference is practically very small, and both effects remain fairly negligible in size (just on opposite sides of zero).

Similarly, looking at Table 4, the correlations among variables are nearly identical across shift and day workers, so the conclusion that there is a relationship between subjective health and behaviors only for male day workers is not valid. (Also, these p values appear to be incorrect? A correlation of -.21 with a sample size of 348 (female shift works) should have a confidence interval of [-.31, -.11] and a p value < .0001, not p = .101 as reported.)

Throughout the paper, please carefully evaluate whether differences across genders or other groups are statistically and practically significant to avoid erroneous suggestions that a relationship is present in one gender but not in the other. A brief examination of results in Table 2 suggests few to no major differences in effect sizes across genders. The authors provide no justification for gender-stratified analyses, and I see no reason to have expected relationships to differ across groups, so I recommend reporting results only for the full gender-pooled sample.
3. For the group comparisons of mean differences (e.g., across shift work groups), please report and interpret effect sizes (mean differences and confidence intervals, Cohen’s d values and confidence intervals) in addition or in lieu of the t test results.

Reporting/presentation of results
4. For transparency and reproducibility of your analyses by readers, please provide a table of means and standard deviations for each variable and a full correlation matrix among all variables. Include the multiplicative composites for all of your interaction variables (e.g., shift work x fluid intelligence). These multiplicative variables are needed for readers to be able to reproduce your results from the correlation matrix.
5. In your regression tables, for clarity, please include step 1 results in the table, as well as in the text. At first, I had thought that Step 1 was a control variables step.
6. To aid readability, in your rgression tables, place the R^2 and ∆R^2 results directly as rows in the table, rather than in the table note. Additionally, I recommend reporting R and √(∆R^2) instead of their squared values to aid interpretability (Funder & Ozer, 2019, https://doi.org/10.1177/2515245919847202). Please also provide confidence intervals for R and √(∆R^2) (or R^2 and ∆R^2), e.g., using the methods described by (Alf & Graf, 1999, https://doi.org/10.1037/1082-989x.4.1.70) or (Shieh & Kung, 2007, https://doi.org/10.3758/bf03192963).
7. In your regression tables, the confidence interval appears to be for b, not Beta. Please place it after SE b for clarity. Please also provide a confidence interval for Beta, using the method described by (Jones & Waller, 2013, https://doi.org/10/gckfx4)

Minor comments
8. (p. 12) A sentence or footnote briefly describing the SOEP study as a nationally representative survey of socioeconomic, health, and psychological variables in Germany would be useful for readers unfamiliar with these data.
9. To help ensure that Zotero is able to keep your citations update to date as you write and potentially change citation styles, I suggest entering prefixes like „for a detailed overview of this project, see” here into the “Prefix” field in the Zotero citation window, rather than directly typing into your document. See https://www.zotero.org/support/word\_processor\_plugin\_usage#customizing\_cites for details.

I hope you find these comments as you revise your work.

Best regards,
Brenton Wiernik

2) Figures/tables/data availability:
Please comment on the author’s use of tables, charts, figures, ifrelevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here).:
The tables are in many places unclearly laid out. No correlation matrix is provided, so the analyses cannot be reproduced.

3) Ethical approval:
If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared.
If not, please detail where you think a further ethics approval/statement/follow-up is required.:
The study uses a large publicly available dataset (https://www.twin-life.de/publikationen/). No information about ethical approval is provided in this paper.

4) Language:
Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.:
The writing is clear.

------------------------------------------------------

------------------------------------------------------
Reviewer B:

1) General comments and summary of recommendation
Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:
 - Are the methodologies used appropriate?
 - Are any methodological weaknesses addressed?
 - Is all statistical analysis sound?
 - Does the conclusion (if present) reflect the argument, is it supported by data/facts?
 - Is the article logically structured, succinct, and does the argument flow coherently?
 - Are the references adequate and appropriate?:
Apologies for the delay in review. I hope the authors are safe and well during this time.

1. I would like to commend the authors’ thorough review of shift work and health behaviors. I learned a lot about this topic from this reading.
2. My general suggestion is for the authors to take a research synthesis approach to summarizing past research instead of individually listing the study results. It might be a personal stylistic preference, but unless there is a strong rationale, I am not sure if statistical results of individual studies such as sample size and odds ratios are necessary to be included. Overall, I think the authors made a good case that the existing evidence is mixed. But I think their presentation can be improved to make the paper a more enjoyable read.
3. I think individual differences cognitive ability as a moderator is an interesting one. However, I am not completely sold by the hypothesizing. Intuitively, it seems to make sense that maybe people with higher cognitive ability use healthier coping strategies. But is this based on theory? Is there past research that support this prediction in other contexts? I would like to see a stronger argument made for this hypothesis.
4. I appreciate the power analysis. Could the authors also report the power of the moderation hypotheses?
5. I realize the authors are interested in gender differences. From a presentation perspective, gender is another variable in the study. Treating male and female respondents as different ‘subsamples’ seems unnecessary.
6. The internal consistencies of the moderator measures are concerning. Are the Cronbach alphas listed on page 13 for the subtests of fluid intelligence? The authors summed the subtests scores – what was the overall alpha for the summed scale? Internal consistencies for the self-control measure is also a bit troubling. How many items were there?
7. The authors used SES as a ‘indicator of cognitive ability’ – is this theoretically appropriate? What is the convergence between SES and actual cognitive ability in the sample? Given the authors have an actual measure of cognitive ability, I am not sure why the authors need to use an indicator like SES. Relatedly, the authors are using SES and ‘education’ interchangeably. I am not sure of SES is the same as education. Instead of equating SES with intelligence, why not discuss how SES might be uniquely related to shift work and health behaviors? I think that is an interesting question on its own merit.
8. The authors analyze the data separately for male vs. female. Why not use a combined sample and use gender as a predictor in the regression equation and then check for moderation or simple slopes separated by gender? The authors’ sample is adequately powered for this. The authors moderation hypotheses do not seem to be contingent on gender, so separating the data into male vs. female seem to unnecessary.
9. In addition to significance tests, the authors should present the difference in shift vs. non-shift work in more meaningful metrics. Given that the variable is dichotomized, I suggest the authors present raw and standardized differences in health behaviors (e.g., Cohen’s d) between shift vs. non-shift work groups.
10. Results on shift work and subj health: please report effect size information (cohen’s d). And again, I am not sure if splitting the samples on gender make sense here. I recommend analyzing the whole sample, and maybe look at gender as a moderator.
11. Occupational status was first mentioned on page 24. How is this variable measured? Table 5 did not give much clarity. The authors noted that they conducted “mean difference analysis” for table 5. I am not sure what that means. The authors report a f statistic – is this in the context of regression?
12. In sum – the authors are examining an interesting topic in a large dataset. I think some of the writing and the analyses could be made much more efficient. I feel the paper could be shortened by 25% without losing too much content and still maintain a coherent presentation of findings.

2) Figures/tables/data availability:
Please comment on the author’s use of tables, charts, figures, ifrelevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here).:
they are adequate

3) Ethical approval:
If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared.
If not, please detail where you think a further ethics approval/statement/follow-up is required.:
na

4) Language:
Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.:
yes

***See a separate supplemental file for the author’s first response letter***

**October 29, 2020**

**Editor second decision**

October 30, 2020

Dear Mrs Myriam Annabelle Baum,

Thank you for submitting your work for further consideration at Collabra. I read the revision and letter of response. I think this is an improved manuscript and well on the way toward something acceptable at this outlet. Thus, I have decided to conditionally accept this manuscript pending your consideration of four remaining concerns. You may dismiss these observations or decide they have merit. Please consider them in the revised final version or in the letter of response.

1. I worried a bit about whether the standards for interpreting effect sizes are almost too stringent in this version. I think talking about the Funder and Ozer framework is useful but I was just a bit worried that the cut-offs were being rigidly applied. In some respects, I think the nature of the predictor and outcome variables should factor into any interpretation and perhaps comparing effect sizes with others in the literature predicting health outcomes is useful. I think it is also useful to think about what happens from the zero-order association of shift work with outcomes to the multivariate association. How did the coefficients for shift work change when considering covariates and is this meaningful? At a minimum, I think you could spell out the approach you will take for interpreting effect sizes in the paper a little more explicitly and make sure this is consistent with the power analysis where a small effect size was used for computations. Perhaps a small effect is meaningful given that these are multiply determined outcomes?

2. Another angle on this issue is to consider the “main” associations of self-control, SES, and fluid intelligence in a little more detail. I think the paper could be enhanced by including just a little more on those variables and some interpretation of how the current results might advance what is known about SES, self-control, and fluid intelligence as correlates of health behaviors.

3. Would it make sense to include a supplement that has SES, self-control, and fluid intelligence in the same model? I think those results might be interesting. [I apologize if I missed this.]

4. As a minor issue – can you add effect sizes to table S5?

To access your submission account, follow the below instructions:
1) login to the journal webpage with username and password
2) click on the submission title
3) click 'Review' menu option
4) download Reviewed file and make revisions based on review feedback
5) upload the edited file
6) Click the 'notify editor' icon and email the confirmation of re-submission and any relevant comments to the journal.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing;, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please do contact us. Please reach out if you have questions - my email is donnel59@msu.edu.

Kind regards,

M. Brent Donnellan
Michigan State University
donnel59@msu.edu

**Author response letter**

Dec 4, 2020

Dear Prof. Dr. Brent Donnellan,

thank you very much for conditionally accepting the manuscript entitled “Reexamining the relationship between shift work and health behavior: Do fluid intelligence, socio-economic status, and self-control moderate the relation?”. Please find the revised version of the manuscript attached. In the following, we would like to briefly comment on the revision we undertook in response to your latest review. Again, in the manuscript, all changes we have made in this version are highlighted in grey.

▪ In the new version, we have relativized the Funder and Ozer-framework (p. 5-6). Along with this, we have revised the presentation of results with a less stringent focus on the Funder and Ozer-cut-offs (p. 34).

▪ We now provide more detailed information about the results (including effect sizes) from contemporary psychological science when discussing our results (p. 34-35). In this vein, we further point out that small effects can be quite meaningful – especially in the context of health (p. 35).

▪ We have added a short section on the main effects of fluid intelligence / socio-economic status / self-control (p. 36-37). However, since this is not the main focus of our work, we have limited this section to select key points.

▪ We now provide supplement material with analyses including fluid intelligence, socio-economic status, and self-control in the same model (standardized and unstandardized; p. 60-65).

▪ We have added effect sizes to the tables S7 (former S5, p.66) and S8 (former S6, p.67).

We would like to thank you for your helpful comments, from which the manuscript has benefited greatly once again.

Yours sincerely,

Myriam A. Baum (for all authors)

**Editor final decision**

Feb 4, 2021

Dear Mrs Myriam Annabelle Baum,

Thanks for addressing the last issues on your submission to Collabra: Psychology, "Reexamining the relationship between shift work and health behavior: Do cognitive ability and personality moderate the relation?". I am happy to accept your submission for publication, pending the completion of copyediting and formatting processes.

I apologize for the delay with this decision. We switched to a new manuscript management system and I missed this paper. The new system is much better so this will not happen in the future. Please do not let this mistake on my part deter you from submitting to this outlet.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point. The accepted submission will now undergo final copyediting. You will be contacted once this is complete to answer any queries that may have arisen during copyediting and to allow a final chance to edit the files prior to typesetting. If you wish to view your submission during this time, you can log in via the journal website.

Congratulations on writing an interesting paper. I hope you continue to think about Collabra:Psychology as an outlet for your work.

Kind regards,

M. Brent Donnellan
Michigan State University
donnel59@msu.edu