Reply to decision letter reviews

https://doi.org/10.31234/osf.io/4br6a

Summary of changes

Below we provide a table with a summary of the main changes to the manuscript and our response to the editor and reviewers:

Section	Actions taken in the current manuscript
General	G + R1:We Added Brenton Wiernik as co-author.
Introduction	ED + R2: We modified table 1 page 5, and page 8-9 to be more explicit about group creativity
Methods	R2 : we updated our search term section G : we specified the databases p 16
Results	
Discussion	
Tables Figures	
Supplementary Materials	

Note. Ed = Editor, R1/R2 = Reviewer 1/2, G = General

Round #1

by Julia M. Rohrer, 25 Oct 2021 07:26
Manuscript: https://doi.org/10.31234/osf.io/4br6a

Request for Revision: Personal factors and group creativity characteristics: A correlational meta-analysis

Dear Dr. Fillon,
I have now received two reviews of your Stage 1 RR, one by a meta-analysis expert and one by a group creativity expert. Based on their feedback, I would like to invite you to submit a revised version of the manuscript that takes into account their central points. While the manuscript shows promise (both reviewers and I agree on that), some clarifications and adjustments will be necessary.
The group creativity reviewer raises some crucial concerns regarding measurement/operationalization and the distinction between the individual and group level. He also provides helpful pointers to the literature. I agree that in the current version of the manuscript, the group/individual distinction is quite unclear, but it is crucial for the present meta-analysis.
The meta-analysis expert, Evan Carter, makes some suggestions for how the methods could be strengthed. He also raises some concerns regarding the role of the sensitivity analyses/how you plan to address publication bias. Considering that this is one of the central concerns in meta-analyses, I suggest you follow up on his idea to implement an approach that produces corrected estimates. None of these approaches are perfect, but I think that the work by Carter is a good starting point to figure out what is most appropriate here.
All points raised be the reviewers seem quite sensible to me. Thus, if there are any that you do not intend to implement, please provide a brief justification/rebuttal in your point-by-point reply.
Best regards,
Julia Rohrer

Thank you very much. For creativity, we changed the part of the manuscript related to it to make sure we only talk about group creativity and not individual level. In the method section, we welcome Brenton Wiernik as co-author, who will help on this subject. He wrote the answer to Evan Carter regarding publication bias/sensitivity analysis.

Reviews
Reviewed by Evan Carter, 19 Aug 2021 19:30
In general, the proposed approach is careful, clear, and well thought out (much as one might expect given the authors' interest in pre-reistration). I was especially excited to

see the authors commit to tracking down unpublished data, as this is one of the most tedious but important steps.

Thank you for this kind comment.

My own expertise is not in psychometric meta-analysis, but looking over documentation for the R package the authors plan to use, I am confident that the analysis will be carried out correctly. Unfortunately, I am not aware of the exact ways in which psychometric MA interacts with publication bias, which I do consider to be my area of expertise. It is my understanding that the primary issue is that low reliability may correlate with publication status or sample size and, therefore, one might incorrectly conclude that bias exists when using typical correction methods. An obvious response to this would be to report on this correlation, which I hope the authors will do for any meta-analytic dataset they are interested in.
A further exploratory analysis could also be proposed in which reliability-corrected effect sizes are analyzed (as in Fig 3 here doi:10.1177/2515245919885611) using standard publication bias correction methods and the kind of sensitivity analysis I have recommended in "Correcting for bias in psychology: A comparison of meta-analytic methods." The issue here, of course, is that this kind of analysis is, to my knowledge, not studied in simulation. That would make it difficult to draw strong conclusions in the event results were difficult to interpret. This exercise might still be very useful.
On a similar point, the authors write, "We also conducted a sensitivity analysis (Mathur & VanderWeele, 2020) with the use of cumulative meta-analysis." I wasn't familiar with Mathur and Vander Weele's work, but in looking over the paper, the authors' sentence doesn't immediately make it clear how they will deal with publication bias via sensitivity analysis. I believe they're referring to section 4.1, and if so, this approach doesn't seem to provide a corrected meta-analytic method, but a sense of whether or not the true effect could be zero and simply inflated by publication bias. In my own work, I prefer to focus on producing corrected estimates as I beleive they make for more useful and impactful meta-analyses. However, if the authors feel that this method meets their needs and will provide useful info for future researchers, I am completely in support of its use here.

As no post-hoc statistical method for detecting and estimating the magnitude of publication bias is perfectly reliable, we will examine potential for publication bias using a combination of approaches (Carter et al., 2019). First, we will estimate PET-PEESE models and examine contour-enhanced funnel plots. These models estimate funnel plot asymmetry, particularly apparent censoring on the statistical significance boundary. Second, we will fit a p-curve to the focal hypothesis test for each study included in the meta-analysis. This method compares the distribution of p values observed to the distribution that would be expected under a null hypothesis (uniform) or true non-zero effect (right-skewed). If the focal p value distribution is closer to uniform or left-skewed than to right-skewed, this indicates against evidential value of a non-zero effect. Third, we will consider whether the effect sizes extracted for each study are the focal hypothesis for the study (as indicated by its presence in the abstract or hypotheses) or an incidentally-reported effect. Publication bias principally affects focal effect sizes, so is generally not a concern for effect sizes that are incidental to the study (Mathur & VanderWeele, 2020). If most of the effect sizes considered are nonfocal, this would reduce the risk of serious distortion by publication bias. The above methods for publication bias estimation have not generally been considered in the context of psychometric meta-analyses that apply statistical corrections for measurement error and/or selection bias (Wiernik & Dahlke, 2019). Statistical artefacts and publication bias have complex interactions. Generally, censoring in reporting and publication focuses on observed, uncorrected effect sizes, so analyses such as p curve and contour enhanced funnel plots should generally focus on observed effect sizes. However, if artefacts are correlated with sample size (e.g., if larger studies have less reliable measures), then publication censoring may be attributable to artefacts such as poor measurement rather than significance per se (Wiernik & Dahlke, 2019). Accordingly, we will construct funnel plots and PET-PEESE models using both observed effect sizes and effect sizes corrected individually for artefacts (for studies with missing artefacts, we will impute artefacts using 2000 bootstrap replicate samples). If apparent asymmetry disappears after correction for artefacts, this may suggest that the asymmetry does not reflect problematic publication bias in the corrected effect sizes.

ref : https://pubmed.ncbi.nlm.nih.gov/33108053/ and https://journals.sagepub.com/doi/pdf/10.1177/2515245919847196

I noticed two other points on which I think clarification would be useful:
1. When results are only in the form of regression coefficients, how will the authors deal with multiple regression models? From what I can see in the supplement, the regression coefficients for which there is a plan seem to come from single-predictor models. The paper, "Concealed correlations meta-analysis: A new method for synthesizing standardized regression coefficients" may be a good resource.

If predictor intercorrelations are available, we can derive the exact correlations, otherwise we will need to contact the author to have access to it. If not available, we will have to exclude this effect size. Regarding the paper mentioned by the reviewer, while being highly interesting, we tend to think that there will rarely be sufficient information in this literature to apply the model proposed. Having direct access to the data or the direct coefficient would be better in any case.

2. Will multiple coders be used per retrieved study? Will inter-coder reliability be reported?
This is an ambitious project and I really think the authors should be commended for their rigorous approach!

Based on another meta-analysis we conducted in personality traits (Lim et al., 2020, IPA), we don’t think we need multiple coders and inter-coder reliability for this meta-analysis. Based on our experience, it is better to have one coder and one independent verifier. We have a column for transparently resolving discrepancies between the main coder and the person who verifies.

Evan Carter
We wanted to thank again Evan Carter for his in depth-review of our dataset and r code and hope we clarified what we want to do and how we will treat publication bias in our meta-analysis.

Reviewed by Greg Feist, 24 Oct 2021 21:19

The proposed research is an interesting and important topic (personal factors and their impact on group creativity). Possible moderators are well-described and many aspects of their meta-analytic procedures are well done and clear. The theory is laid out well and explicitly.

Thank you for this comment.

There are serious problems with the proposal, however. Replication would be difficult because there are problems of confusion/clarity mostly dealing with measurement/operationalization issues. The biggest problem I see in this proposal is the authors are very unclear on how to operationalize the key components of the study, namely team creativity and personal factors of the individuals within the teams. The obvious problem is measuring individual personality and cognitive data per individual team member and then measuring creativity at the group level. The researchers certainly have an answer to this but, as far as I can see, they do not make that clear anywhere in this proposal. As someone very familiar with personality and creativity research I need them to spell out the mechanics of individual versus team level measurements. For example, let's say a team has 10 people. That is an N of 10 on personal factors. But team creativity (fluency, originality, etc) of the team and has an N of 1 (one team). So how do you correlate across levels of analysis?

Indeed, we might not have been clear enough. This meta-analysis purpose is only at the group level. There are plenty of studies and meta-analyses at the individual level but to our knowledge, not a lot at the group level. We based our meta-analysis on Coursey et al. (2018) review on the relationship between group creativity and personal factors. This gap between global review and lack of clear meta-analysis is the main reason for this meta-analysis, we want exclusively to see how personality factors interfere with the group creativity.

Similarly, in their Introduction they confuse/conflate research at the individual and group levels. For instance, Table 1 title says "personal factors in creative groups" but then the researchers review studies that do not deal with group creativity (e.g. Furnham, Batey, King, etc). Other studies in the table are clearly at the group level (Bechtold et a., 2010). So these seem to be confounded in this table. Similarly, in the section on "Relationships between Personal Factors and Creative Activity Characteristics" (1st paragraph), they say there is a debate regarding relationships between personal constructs and group creativity and then cite research that was not at all group-based (e.g., Feist, 1998).

This is 100% correct. We completely changed Table 1 to make sure studies are at group level. We modified the section below the table, by making explicit the difference between individual and group levels.

The lack of clarity about group continues in the Design section (p. 8). Creativity outcomes are described as "number of ideas generated," "originality of these ideas,"

and “usefulness of the ideas” without specifying group and without operationalizing originality and usefulness.
Other scholars have been more clear about these problems. As Litchfeld et al (2017) discuss in their chapter, team personality (and creativity) must spell out whether it is measured via the composition or compilation method. The current proposal does neither. Even so, in my mind, the composition method has a problem, since it derives a team-level score from either the mean or variance of that trait. The mean without variance within a group can be very misleading. A mean of 50, for example, could be derived from a team that varies little or a lot around that mean of 50. Yet, substantively, group with a little or a lot of variance are different groups. The variance without the mean is better since you can have high and low heterogeneity groups on a personality dimension. But this distinction is not made in the proposal.

Coursey et al (2010) also discuss the dynamic and potentially synergistic effects of individuals working in groups and distinguish between the additive, contingent, and configural approaches. And they also explicitly discuss “aggregate Openness” for instance when discussing personality and group creativity. This kind of discussion is missing in the proposed study and is needed to clarify how person factors and group creativity are operationalized.
The same problems that exist in operationalizing personality also exist for team creativity. But the authors never even address this question. Is it team creativity via composition or compilation method? Does the team get one score or many on each creativity outcome (e.g., originality, fluency)?

We only wanted to take the correlation found by researchers for the purpose of the meta-analysis, and report in another column how the construct was measured so that the way researchers measured the construct was transparent.
Concerning the variance, we will take the reliability coefficient into account for the correlation, and the reliability coefficient is created based on the variance of the total score for all participants. If the reliability coefficient is not available (which is rare), we will use the artefact distribution to input a coefficient. After that, for the meta-analysis, we will weight each effect size based on the sample size.
In other words, in our meta-analysis, we will use the “compilation method” with the use of the correlation between the trait and the creative outcome at the group level, corrected with the use of both reliability coefficients which contain the variance of individuals composing the groups. We didn’t used the specific term of “compilation method” since it is not common in psychometric meta-analyses, but we took that into account.

For the most part, the hypotheses are meaningful. Publication status, however, is bit obvious and well-established in the meta-analytic literature, namely the larger effect for published versus unpublished studies. I am not sure that adds anything to the study or the literature.

Thank you, the published vs. unpublished studies is a typical way of addressing possible publication bias. If the published studies have reported a stronger association than unpublished studies, it would mean that the effect is inflated in publication, and we need to be more cautious in reporting it in our meta-analysis. This is more a moderator about the “confidence” we can have in the effect size we will find and is not specific to this literature.

In keywords for the literature search, I don't see "brainstorming." They say in their Intro that is synonymous with group creativity, so it is surprising that is not included. The authors do a good job of avoiding problems associated with Null Hypothesis Significance Testing (NHST), but met-analyses generally do. I am not sure, however, there is need to use their "meaningful" criterion of r > .10 since effect sizes have their own more established "rules of thumb" for small, medium, and large effects (see Cohen). But this is not a critical issue.

Thank you for this comment. Indeed, we will include brainstorming and brainwriting. the exact pattern for overall creativity is: "creativ*" OR "idea generation" OR "problem solving" OR "brains" OR "brainw" AND "group*" OR "collab*" OR "team". We modified the main manuscript page 17 accordingly. We also wanted to add that all research patterns are transparently reported in the spreadsheet in the sheet "search pattern".
About the effect size, we decided to draw our conclusions based on this criterion of r = .10 for the discussion section. Other researchers might use the exact effect size or a specific effect size based on their knowledge and use it for their own purpose (as, for example, a planification for sample size).

As a meta-analysis, the ethical issues are minimal to none and there is no untoward conflicts or problems with this study. No IRB is required.

Thank you very much for your careful examination of the introduction, design and operationalization of our study. With your help, we highly improved the scope of the study.