General reply

Before responding to reviewers' specific comments, we would like to make a few points.

The purpose and hypotheses have been greatly simplified. We are investigating the relationship between creativity and trait depression, specifically testing whether the relationship becomes negative or changes in a negative direction, when two factors are controlled: rumination and reappraisal frequency.

This simplification addresses some reviewer comments that the model was convoluted and possibly not falsifiable. We now have three simple hypotheses that can be addressed with multiple regression.

Previous literature shows mixed evidence regarding the relationship between creativity and depression, and no studies have tested whether the relationship between creativity and depression would change when these emotion regulation strategies (ERSs) are controlled. The only relevant study would be the one by Verhaeghen et al. (2005), who tested self-reflective rumination as a confounding variable in a complicated path model. Our present study draws upon their model in a simplified way and tested one more ERS of reappraisal.

One difficulty is that since no studies have tested the same hypotheses, we do not have a good prior estimate of effect sizes, and there is a practical limit to how small of an effect was can target. We assessed how many participants we could recruit given our available funds and the fact that the task requires about 13-15 mins (we have shortened our creative tasks because the long duration could deter subjects from participating), and the maximum sample size is about N=200.

We think that our study could contribute to the literature even if we do not have enough power to detect small effects. A sample size of N=200 could detect small to medium sized effects, or could show trends that could be used as the basis for further study. The width of the 95% confidence intervals for correlation estimates would be between .25-.28 for small correlations (r=0-0.3), and .21-.25 for medium correlations (r=0.3-0.6). This would provide some information even if effects are too small to detect.

An advantage of the RR format is that results would be published regardless of whether statistically significant effects are observed. The measures resulting from our study would provide unbiased estimates of the overall relationship and the effect of controlling the two factors.

Reviewer 1 Comments

Overall, I think the rationale for the research questions could be much tighter and more logical and clearly explain why several of the hypotheses seem counter-intuitive at a first glance.

We have performed several major changes to the introduction:

1. We skipped over the association between creativity and other mental illnesses, and only described the linkage between creativity and depression
2. We provided a brief summary of our arguments, which mentioned that the present literature shows inconsistent findings regarding how creativity and depression could be linked, and the theoretical framework demonstrates that these two variables are unlikely to be related directly. This raises further questions about how creativity and depression could interact through other pathways.
3. We adapted the section about reappraisal frequency and revised our hypothesis. As no studies have investigated the relations between reappraisal frequency and creativity, there is not enough evidence for us to predict that it has a mediating effect between creativity and depression. Instead, we hypothesize that reappraisal frequency to be a suppressing variable which allows us to more openly observe how it influences the relationship between creativity and depression. More details can be found under "Emotion regulation strategy: Reappraisal" in the introduction.

The rationale as it stands fee convoluted. While the broad premise is that mental illness and creativity are anecdotally linked, many of the examples are more related to bipolar or schizophrenia where the mania element might be more logically linked to creativity. Focusing more quickly on why they are looking at depression seems warranted given that it is characterised by problems of motivation, concentration, and poor executive function, all of which do not seem particularly conducive to creativity, and as the authors acknowledge there is good evidence to suggest that creativity might be reduced in depression – especially during an acute episode – which undermines the rationale for the study somewhat.

We revised the introduction to focus more specifically on depression.

We skipped over the broad premise that mental illness and creativity are anecdotally linked, and quickly focused on the idea that depression could be related to creativity. We corrected the historical examples of famous people who were diagnosed with depression.

We also introduced the existing problems in creativity research: (1) the present literature also shows inconsistent findings and (2) the empirical evidence appears to be a contradiction with the theoretical understanding of creativity and depression. The rationale of our investigation is based on these research gaps.

Reappraisal ability has been linked to creativity, but the authors are interested in reappraisal frequency, suggesting that ability is not necessarily impaired in depression, but people just reappraise less frequently. Can they expand on how this relates to creativity? Might this also imply that people with depression can be creative, but they engage in creativity less frequently? These seem like an important nuance that isn't currently captured.

We have revised our hypothesis. We hypothesize that reappraisal, just like rumination, influences the association between creativity and depression. Specifically, we hypothesize that when we control for reappraisal frequency, a negative association between creativity and depression may emerge. Research shows that creativity is positively related to reappraisal ability, which could reduce depression in theory. But, studies showed that even when people have a high reappraisal ability, if they did not use reappraisal frequently, this will still contribute to higher levels of depression. We hypothesize that creative people may fall into such descriptions – having a low frequency may mask the potentially beneficial effects of reappraisal ability, which is related to higher depression.

With our revised hypothesis, we will test and replicate the effects of rumination first according to Verhaeghen et al. (2005), which is that self-reflective rumination is a confounding factor for creativity and depression and there is no direct relationship between creativity and depression. Then, we will add reappraisal frequency to our existing model. We hypothesize that when reappraisal frequency is controlled, we may even observe a negative association between creativity and depression.

We revised our discussion of reappraisal ability and frequency to make the logic clearer.

Finally, the clinical implications could be more sophisticated – CBT already targets rumination and promotes reappraisal – how will understanding whether these are related to creativity change this?

We removed discussion of the clinical implications to reduce confusion.

The protocol describes a simple survey containing several questionnaires. The inclusion criteria for participants seem very broad – all subjects at least 18 years old. Have the authors considered whether they wish to include people with other mental health diagnoses given the discussion in the introduction? Depression is also skewed in the population with most people reporting no or few symptoms and a tail with increasing severity. Given that depression is the main outcome of interest the authors could consider a more efficient sample strategy, over recruiting people with more extreme scores to ensure they recruit an informative sample. Otherwise, they will end up with lots of people with few to mild symptoms which will reduce their ability to detect effects. It also seems important to assess whether people are currently undergoing treatment and whether this should be an exclusion criterion or not – we know that both antidepressants and CBT affect cognitive processing so may influence measures of creativity, rumination, and reappraisal. It also seems remiss to not capture other important demographic information like profession, given that previous

work has looked at whether rates of depression differ between creative art type professions and others.

We have updated the following exclusion criteria in the 'Participants' section:

1. Subjects who have experience with Torrance Tests of Creative Thinking or any other creativity thinking tasks
2. Subjects with bipolar disorders or schizophrenia, because these these diagnoses are associated with the component of mania or positive symptoms, which are found to be associated with enhanced creativity (e.g. MacCabe, 2018; Power et al., 2015; Silvia & Kimbrel, 2010) and also higher depressive symptoms (Bosanac & Castle, 2013; McCormick et al., 2015; Stamouli, 2010; Upthegrove et al., 2017).

We will not exclude subjects based on current or past diagnosis of clinical depression, or based on current or past treatment of depression. Including such people would be useful for measuring the larger variations in depressive symptoms and our assessment of depression, creativity, and ERS.

We also clarified that we aim to measure the variations of depression in the general population. Therefore, we will not recruit a strictly depression sample, but a sample with varying from no experiences with depression to those with severe depressive symptoms.

For the demographic information of "profession", obtaining and analyzing this information might divert our focus from ERSs. As we do not have a specific hypothesis about the interactions among professions, creativity, ERSs and depression, we will not ask participants about this.

The authors are looking at trait depression, but it seems important to also measure current depression and mood as these could be important confounders. Indeed, their analyses do test their hypotheses in a crude way but have the authors considered potential confounding factors like current mood, medication ect.

We chose to measure trait depression instead of state depression because creativity is theorized to be a stable dispositional trait rather than a temporary state (e.g. Feist, 1998; Puryear et al., 2017; Zhang et al., 2020). If creative individuals have greater tendency toward depression, it would be revealed in trait depression as both variables are relatively permanent and consistent. Measuring trait depression also allows us to assess the normal variations of depression in the general population, making this study an epidemiological study rather than a diagnostic one. Hence, we will not inquire about participant's current mood as this may not be related to our investigation of depression as a lifetime experience.

This clarification was added to the section under "Methods > Materials > Depression"

We also decided not to include medication as exclusion criteria, because including such people would be useful for measuring the larger variations of depressive symptoms in the general population, which will be helpful to our study of depression, creativity, and ERS.

It would help to describe what higher scores mean on the measures, and what the range of possible scores is. It would also be good to clarify how the measures will be operationalised as DVs (i.e., sum scores) and whether only sub-sections will be used. For example, the rumination measure has several sub-scales only one of which is related specifically to self-rumination which features in their hypotheses.

We have added more details about the range of possible scores, as well as the operationalization of sum scores/ subscores to the 'Materials' section.

The methods are missing a detailed analysis plan. This would help to describe the main DVs and IV and clearly state how each hypothesis will be tested. This is partially covered in the framing of the results section and in Table 1. The authors could also consider how they might control for / investigate the effect of potential confounding factors mentioned above.

An analysis plan has been added to the 'Method' section. We described the IVs and DVs to be tested for each hypothesis. The outline of the "result" section also shows what numbers will be reported.

Table 1 provides an overview of the research questions, proposed analyses and possible interpretations. There are some logical inconsistencies. For example, a negative association between rumination, depression and creativity is predicted yet the box below suggests the evidence to suggest a positive association.

We have revised our hypotheses. Regarding rumination, we hypothesize that when self-reflective rumination is controlled, the positive associations between creativity and depression will be reduced.

Table 1 explains how their analyses confirm/disconfirm their hypotheses. They are using NHST. This is fine for when the null is rejected. They will not be able to claim evidence of an absence of effect, however, unless they also incorporate Bayesian inferential tests.

In addition to reporting regression coefficients and standard NHST results, we will also compute and report Bayes factors. This has been added to the analysis plan.

The chosen effect size for the sample size justification is relatively well justified. However, it seems a bit optimistic given that it is towards the upper end of several of the ranges of previously reported effect sizes, and we know that many published effect estimates are inflated by publication bias. Given this is an online survey with broad inclusion criteria and data collection is relatively easy, the sample size could easily be increased to ensure the authors don't miss effects. I would suggest being conservative and powering for the smallest likely effect of interest.

The maximum possible sample size is limited to N=200, due to funding and time constraints. We think this could be informative even if we cannot guarantee power to detect small effects.

This is broadly achieved. There are just a few places where the chosen effect of interest r>.26 is toward the upper range of previously reported effect estimates ((e.g., r = .09-.35; the correlations observed in individual studies ranged from r = -.14 to -.29 and the overall correlation estimated across studies was r = -.17)

The correlations of r=.09-.35 by Verhaeghen et al. (2005) were between different creativity measures (including fluency, originality, elaboration) and self-reflective rumination in a large and complicated path model. The correlation between the total creativity score (by adding up fluency, originality, elaboration) and rumination would likely be larger than the independent path between each creativity measure and rumination.

Our current target sample size of N=200 could have 80% power to detect small effects of r>.19, or 95% power to detect medium effects of r>.25.

Have the authors avoided the common pitfall of relying on conventional null hypothesis significance testing to conclude evidence of absence from null results? Where the authors intend to interpret a negative result as evidence that an effect is absent, have authors proposed an inferential method that is capable of drawing such a conclusion, such as Bayesian hypothesis testing or frequentist equivalence testing?

Table 1 suggests NHST will be used and that they will be able to find evidence for or against current theories. To achieve the latter, they will need to add additional analyses.

We will perform Bayesian analyses and 95% CIs of the estimated posteriors for the effects. This will provide some information about the strength of evidence if we observe non-significant trends.

The authors have not included an analysis plan in the methods and Table 1 discussed only planned analyses. In the results section the authors state that if creativity is found to be associated with reappraisal, rumination or depression, they will conduct additional exploratory analyses using separate measures of creativity: fluency, flexibility, originality

We have added an analysis plan to the 'Method section', including exploratory analyses.

In short, we will conduct the following analyses:

1. Hypothesis 1: linear regression with creativity (IV) and depression (DV).
2. Hypothesis 2: add self-reflective rumination (IV) to the multiple regression of creativity (IV) and depression (DV)
3. Hypothesis 3: add reappraisal frequency (IV) to the multiple regression of creativity (IV), self-reflective rumination (IV) and depression (DV)
4. Exploratory analyses:
    a. If creativity is found to be associated with reappraisal, rumination or depression, we will conduct additional exploratory analyses using separate measures of creativity: fluency, flexibility, originality. We will perform a correlation matrix using the variables of fluency, flexibility, originality, reappraisal frequency, self-reflective rumination, and depression.

5. We will use rumination (IV), gender (moderator), and depression (DV) in the first moderation analysis, and reappraisal frequency (IV), gender (moderator), and depression (DV) in the second moderation analysis.
6. Inferential tests: Bayesian Linear Regression, using creativity, self-reflective rumination and reappraisal frequency as IVs and depression as DV

More details can be found in the analysis plan.

There are no data quality checks reported. The authors could include attention checks in their survey and describe any exclusion criteria.

We have added attention checks to the survey (more details in 'Materials'). An example of an attention check "This is an attention check. Please select "strongly disagree" for this question." Subjects who failed to answer them according to instructions were excluded in the analysis.

As mentioned, the exclusion criteria has been updated in the 'Participants' section:

1. Subjects who have experience with Torrance Tests of Creative Thinking or any other creativity thinking tasks
2. Subjects with mental illness diagnoses of bipolar disorders and schizophrenia were also excluded, because these diagnoses are associated with the component of mania or positive symptoms, which are found to be associated with enhanced creativity (e.g. MacCabe, 2018; Power et al., 2015; Silvia & Kimbrel, 2010) and also higher depressive symptoms (Bosanac & Castle, 2013; McCormick et al., 2015; Stamouli, 2010; Upthegrove et al., 2017).

The authors do not seem to mention ethical approval. Adding some comments on the ethical considerations of the study seem warranted – especially given the population of interest.

We have received ethics approval from the Departmental Research Ethics Committee for the current study. Details about ethics have been added to the 'Procedure' section.

Reviewer 2 Comments

The research question on the link between creativity and depression as well possible mediation through reappraisal and rumination is well derived. I agree that there is still a lot of discrepancy on whether creativity is associated with more or less depressive symptoms and real-life emotion regulation tendencies may play a major part in this relationship. The research questions proposed by this paper are definitely valid; however, I am not convinced that the authors did a thorough enough literature research to back up their research questions and hypotheses. Especially in the "Creativity" and the "Creativity and depression" paragraph, I am missing critical references for a) the categorization of creativity, b) the notion that divergent thinking leads to originality, c) that divergent thinking abilities are indicative of creative thinking and creative potential, and d) that anhedonia inhibits creativity. Adding more references would in my opinion boost the credibility of the research proposal.

The following references have been added:

a) the categorization of creativity: Others have categorized creativity into domains (e.g. everyday, visual, verbal, performance, scientific etc.) (e.g. Taylor, 2017; Villanova & Cunha, 2020), achievements (e.g. Carson et al., 2005), professions (e.g. Ludwig, 1992) or activities (e.g. painting, writing, musical composition etc.) (e.g. Hocevar, D. 1980; Verhaeghen et al., 2005).
b) the notion that divergent thinking leads to originality (Runco & Acar, 2012; Kim, 2017)
c) that divergent thinking abilities are indicative of creative thinking and creative potential (Cramond, 2020; Kim, 2017)
d) that anhedonia inhibits creativity (Shapiro & Weisberg, 1999; as cited in Verhaeghen et al., 2005)

We have also revised the reference list accordingly.

1) The authors state that they want to address "the inconsistencies in the current literature that blurs the distinction between reappraisal ability and reappraisal frequency". In my opinion, this is only possible of this study investigates BOTH reappraisal ability and reappraisal frequency as a mediator between creativity and depression. To me, it is problematic that the authors build their hypotheses for reappraisal frequency on the idea that creative individuals have high reappraisal ability, yet there are only a handful of studies so far that propose this link for very specific measures of creativity and reappraisal (e.g., Weber et al., 2014; Fink et al., 2017). Accordingly, I strongly recommend that the authors include a measurement of reappraisal ability in their study as well. I elaborate on possibilities in 1C.

After careful consideration, we decided not to include the measures for reappraisal ability due to several reasons.

Adding the measures of reappraisal ability would take at least 15 mins more for participants to complete the task, which makes the entire questionnaire over half an hour long. This might deter subjects from participating in the study, and would make it too expensive to collect a large sample size.

Furthermore, there is reason to expect that a measure of reappraisal ability would be redundant. The measures for reappraisal ability (such as Reappraisal Generation tasks (RGT), Reappraisal Inventiveness Test (RIT), or the Script-based Reappraisal Test (SRT)) were developed based on divergent thinking activities. Both are also scored according to similar grading criteria of fluency and flexibility, which is the number of ideas generated and how categorically different the ideas are (Wu et al., 2017). Specifically, the RGT and RIT were found to be significantly associated with verbal divergent thinking tasks (Fink et al., 2017; Perchtold et al., 2018), which is the measure we will be using in the current study (TTCT-verbal). These details have been added to the "Creativity and Reappraisal" section.

We expanded our review of reappraisal ability and cite more evidence about the strong associations among reappraisal ability, divergent thinking, and creativity. More studies were included in the discussion of "reappraisal" to illustrate the interconnectedness among these concepts, and revised the wording to make our concept clearer.

Because incorporating reappraisal ability measures would compromise our timing, is secondary to our hypothesis and is related and similar to our creativity measure, we decided not to include it.

In addition, we have addressed and clarified our hypothesis regarding reappraisal frequency.

---

2) Further, another major factor that may influence links between creativity, depression, and emotion regulation strategies is gender. There is a plethora of research discussing gender differences in adaptive and maladaptive emotion regulation strategies like reappraisal and rumination and their link to depression, for example:… I think it is important to consider that according to these publications, women use both adaptive and maladaptive emotion regulation strategies more than men. Simultaneous use of both adaptive and maladaptive strategies may reduce the effect of adaptive strategies (e.g., reappraisal), which could also affect the link between creativity and depression in women. Thus, I think it would help clarify associations between creativity, depression, and emotion regulation if gender as additionally considered in all analyses (see 1C.)

We will consider gender in our exploratory analyses. Some studies have indeed found gender as a moderator between rumination and depression, and between reappraisal frequency and depression. A brief review of literature has also been added to our paper. But, as several studies failed to replicate similar results, the present literature is inconsistent. Gender is also only secondary to our main model. Therefore, we will only examine gender as an exploratory analysis.

We will perform two separate moderation analyses using rumination (IV), gender (moderator), and depression (DV) in the first analysis, and reappraisal frequency (IV), gender (moderator), and depression (DV) in the second analysis. This data could explain some possible variance in our model.

a) Are there any exclusion criteria for study participation? What comes to mind is current psychiatric/neurological diagnosis, intake of psychoactive medication, and previous experience with the TTCT. It is also unclear whether the authors want to investigate a sample with normal variations in depression or whether individuals with high depressive symptoms will be included/excluded from the study.

We have revised the exclusion criteria. We will exclude subjects who have experience with Torrance Tests of Creativity or any other creativity thinking tasks, as well as subjects with bipolar disorders or schizophrenia, because these diagnoses are associated with the component of mania or positive symptoms, which are found to be associated with enhanced creativity (e.g. MacCabe, 2018; Power et al., 2015; Silvia & Kimbrel, 2010) and also higher depressive symptoms (Bosanac & Castle, 2013; McCormick et al., 2015; Stamouli, 2010; Upthegrove et al., 2017).

But we will not exclude anyone who is using any particular psychoactive medication as these medications may be used for various functions or treatments unrelated to depression. We will also not exclude people with medication/ treatment for depression or those with a history of depression, because these subjects can be helpful in adding variations to our sample.

To clarify, we aim to investigate a sample with normal variations of depression in the general population. This includes people who are clinically diagnosed with depression, as well as people with no to mild depressive symptoms. We have clarified this in the main text as well.

b) What is the protocol if participants score very high on the Depression Scale (MSD-T)? The authors mentioned that they could not provide sufficient professional knowledge and advice regarding the CES-D, which leads me to believe that the authors planned to follow-up on individuals with high depression? How will this be accomplished?

Since we are measuring trait depression instead of state depression, this is indicative of normal variations in depression but not a clinical or diagnostic problem. As this is an epidemiological study, we will not provide follow-up on individuals with high depression. We have also received ethical approval for measuring trait depression without clinical follow-up.

c) Will participants be financially compensated for their participation?

Participants will receive financial compensation depending on the allocated funding. We have inserted a placeholder sentence in the main text and will update the amount in the future.

d) There should be more clarity on which verbal subscales of the TTCT will be used. Additionally, how many raters will assess fluency, flexibility, and originality of ideas and how will the interrater-reliability be computed?

Four verbal questions from TTCT will be used. There is no further subcategorization of verbal subscales. Our questionnaire has been uploaded to our preregistration.

Interrater reliability was assessed by having one rater score the TTCT according to the guidelines on the TTCT interpretive manual (Torrance, 2018), then a second rater independently scored a subset of 30 responses. We will perform the intraclass correlation coefficient (ICC) across the items and report the value.

As states above (1C), more detail on the use and scoring of the TTCT should be provided. This also applies to any measure of reappraisal ability/capacity that the authors choose to implement.

More details about the use of TTCT have been added, including new exclusion criteria and interrater reliability.

Details about the scoring of the TTCT were also added:

For fluency, the raters will count the number of answers; for flexibility: raters will count the number of categories; for originality, raters will count how many times an answer is duplicated among other participants – which suggest that the answer is not original. To obtain the originality score, this number is inverted using 1/n. The three sub scores are added to form a total score.

The authors briefly address their protocol for excluding participants from data analysis (blank responses to the creativity tasks, failure to complete all sections, etc.). Given that the study is conducted online via Amazon's Mechanical Turk, I feel like more rigorous data quality control should be applied. For example, if participants are paid for study participation, are there any measures in place that prevent participants from filling in the surveys multiple times? Will TTCT answers be screened for plausibility? Will the overall test taking time be considered? (could point to arbitrary answers on the questionnaires). I would like the authors to elaborate on their security protocol to avoid that data from automated bots is treated at real participant data.

Attention checks were added to ensure that participants read the instructions carefully. Subjects who failed to answer them according to instructions will be excluded in the analysis.

As our creative measures require written responses, performance by bots would be difficult and easily detected through inspection of answers. We will manually screen the TTCT answers for plausibility based on how appropriate or relevant the responses are, and will exclude any nonsensical answers.

As the verbal creativity tasks require written answers and are quite time-consuming, it is also unlikely that participants will file the survey multiple times.

The overall test taking time is controlled by the timer of the TTCT. Participants cannot proceed until the timer is up (2 mins per question, timed 8-10 mins in total). This was added to the main text under "Methods > Materials > Creativity".