Thank you for the opportunity of reviewing this study. There are several areas where improvements could be made. I have made several comments/suggestions below in response to the reviewing guidelines. I hope the authors find them useful.

* Does the research question make sense in light of the theory or applications? Is it clearly defined? Where the proposal includes hypotheses, are the hypotheses capable of answering the research question?

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.

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.

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.

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?

* Is the protocol sufficiently detailed to enable replication by an expert in the field, and to close off sources of undisclosed procedural or analytic flexibility?

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.

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.

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.

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.

* Is there an exact mapping between the theory, hypotheses, sampling plan (e.g. power analysis, where applicable), preregistered statistical tests, and possible interpretations given different outcomes?

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.

* For proposals that test hypotheses, have the authors explained precisely which outcomes will confirm or disconfirm their predictions?

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.

* Is the sample size sufficient to provide informative results?

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.

* Where the proposal involves statistical hypothesis testing, does the sampling plan for each hypothesis propose a realistic and well justified estimate of the effect size?

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)

* 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](https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00781/full) or [frequentist equivalence testing](https://journals.sagepub.com/doi/full/10.1177/2515245918770963)?

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.

* Have the authors minimised all discussion of post hoc exploratory analyses, apart from those that must be explained to justify specific design features? Maintaining this clear distinction at Stage 1 can prevent exploratory analyses at Stage 2 being inadvertently presented as pre-planned.

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

* Have the authors prespecified positive controls, manipulation checks or other data quality checks? If not, have they justified why such tests are either infeasible or unnecessary? Is the design sufficiently well controlled in all other respects?

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

* Does the proposed research fall within established ethical norms for its field? Regardless of whether the study has received ethical approval, have the authors adequately considered any ethical risks of the research?

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.