**\*We wish to thank these 2 anonymous reviewers for their comments and positive response to the paper. We have endeavored to address each point in turn and believe the paper is a stronger one as a result. Our responses are denoted herein with an asterisk and bold text. Thank you again for taking your time to review and provide feedback on this meta-analysis.**

This is an interesting and well-researched paper that, with revisions, could be influential in clinical, counseling, health, and other areas of psychology. Whereas the statistical rigor is evident, the authors would have a far more powerful paper if they rewrote parts of it with their likely audience in mind. Three general issues need to be addressed in a revision:

1. **Clean up the logical and literature review**. The introduction, especially the first 3-4 pages, is not focused and needs significant tightening. The “emotional expression” section relies on correlational and experimental studies that vary in their quality and in their ability to identify a causal relation. There is very little causal evidence, for example, that inhibition or repression causes illness. The literature review under “Expressive writing as effective emotional expression” needs greater structure. The third section on “possible mechanisms underlying WED efficacy” doesn’t have a center. It’s not clear why this section is here since it doesn’t lead up to the point of the paper. In a revision, everything up to the middle of page 6 should be geared to the study you are about to present.

**\*Thank you for bringing these structural issues to our attention. We restructured the emotional expression section to include the following in different paragraphs: an explanation of expressive writing and the psychological/physical outcomes and a correlates of emotional inexpression (operationally defined as cognitive avoidance behaviors) section that also focuses on potential power issues within the expressive writing literature. Additionally, we combined the expressive writing as effective emotional expression with the meta-analytic techniques section to enhance the structure and clarity of our introduction. We excluded the third section on possible mechanisms underlying WED efficacy. As pointed out in the review, this section fails to effectively segue into the current study.**

The meta-analysis section begins to prepare the reader for the actual paper. Remember that your likely readers will be clinicians and health psychologists. Some may be as sophisticated in statistics as you, but most won’t. Your goal is to communicate your ideas and not show off. For example, the different previous meta-analyses have relied on *d*, *r*, Hedges *g*, *g*post, etc. Convert all of these statistics into a common currency (my recommendation is *d*, but your call).

**\*The values in the introduction have been converted to *d* for comparison and clarity.**

A recommendation on writing style: make paragraphs shorter. Any paragraph that goes over a page and that introduces multiple topics should be segmented. This will make the paper far far more readable.

**\*Thank you for your writing style recommendation. We have shortened the paragraphs going over a page to enhance the readability of the paper. Additionally, we have separated paragraphs that introduce multiple topics.**

2. **Rethink the presentation of statistics and results**. Beginning on page 15, the paper becomes increasingly unreadable for readers not intimately familiar with your statistical approach. You are dealing with incredibly important topics that very smart people are interested in. If your goal is to show off your statistical chops, then put pages 15-19 in supplemental material and replace these pages with an English rendition of what your methods do.

**\*This section contains the details pertinent to the meta-analysis, such as relevant formulas, and descriptions of the analyses that would be found in the results section. We have left these sections intact, as there was concern that if we removed them, the results would be uninterpretable, especially to a naïve audience.**

In the Results section, the findings start to become more understandable, although simplifying the language could help at points. The readers likely need more information, however, about the ways you are looking at only individuals in the experimental conditions and how they change over time. By looking at only experimental conditions, the Ns of the studies are generally cut in half. Some of the studies also have multiple experimental conditions – are all included?

**\*Yes, all experimental conditions are included. This information was added to the data collection section.**

How exactly did you calculate the time factor? Was this a within-subject variable or a between-subject. Some of the studies may have measured people 1 day after writing and then again 3 and 6 months later. Others may have just measured people one month after. Say more in early methods and remind us here.

\***All data was within subjects by using only the experimental groups across time. Time between assessments was coded as the number of months between points, and this information was added to the data collection section.**

Although you mention this briefly earlier in the manuscript, expand on the logic of not looking at the control group data. From a research design perspective, most clinicians are interested to know if WED or other method works better than a placebo control. Isn’t it possible that the effects you find simply reflect a reduction in PTS over time that would always be found in ratings of anxiety or stress even for people not in an experiment? QOL measures, on the other hand, reflect positive affect which may not show a similar drop. These issues need to be rehashed in the discussion.

**\*Information on this idea was added to the discussion.**

Walk us through the graphs more carefully. For example, the PTS graphs are all fairly remarkable. For each graph, give us the grand mean of d. In fact, a simple summary table of the d-values would be helpful. Also, we need to know what the experiment initials refer to in the graphs. These should be marked in the references.

**\*The grand mean of the effect sizes are not provided on the graph, as they are provided in Tables 1-3. The values in Tables 1-3 are the weighted meta-analytic effect size (based on analysis type and fixed/random effects), and a simple graph average would not accurately represent the overall effects. The PTS graphs are split by subscale to provide graphs for them, as they would not all fit on one graph. The coding of the Y-axis is provided in the data (on OSF** [**https://osf.io/4mjqt/**](https://osf.io/4mjqt/)**) – these were abbreviated due to size limitations. We added notes to the figure caption to help direct readers.**

**APA style denotes that once meta-analyses have more than 50 articles, a second reference list is created for the articles, which was included with our paper. Therefore, each set of labels is included that second reference section (which we realize may have been excluded in our first submission by mistake).**

3. **Effect sizes and their implications**. The ways effect sizes are discussed in the paper fails to appreciate some fundamental problems in clinical research. Note that the description and interpretation of effect sizes at the top of page 15 certainly holds for lab studies conducted in highly controlled settings where the outcome measure is collected seconds or minutes after an intervention. Many, perhaps most of the expressive writing studies have outcome measures that are measured days, weeks, and even months after the intervention. Indeed, these are comparable to medical studies. In many of the most famous medical studies, effect sizes are routinely under d = .05. This is true for the effectiveness of chemotherapy and/or radiation on cancer; taking aspirin after a first heart attack to reduce the effects of a second attack, etc. It is no wonder that many of the most important medical trials employ 10,000 people. The true effect sizes are minuscule but still important.

It is also not coincidental that WED average effect sizes are in the d= .11 to .18 range. A brief, rather innocuous intervention still has effects weeks and months later. That you are finding much larger effects is truly remarkable. In other words, as even Cohen would argue, an effect of .20 using long term real world measures is probably a medium possibly large effect.

**\*Agreed. Thank you for bringing this to our attention. We have added writing in the discussion section to include a caveat for interpreting these results within the context of the intervention.**

One other issue in the discussion. The cost to run an expressive writing study is surprisingly high in terms of personnel and the time it takes to do such a study. Your observations that most such studies are underpowered is true. By the same token, your recommendation would mean that no expressive writing studies would ever be run in the future. Are you recommending that the entire research enterprise be shut down? I would urge you to think how the field must deal with this very complex problem. Sure, if you are a computer scientist and you are analyzing the data that these hundreds of researchers have collected, you can say any single study is problematic. At the same time, you have done a nice job showing us important and trustworthy findings by relying on our large number of costly and time-intensive underpowered studies.

**\*Thank you for bringing this to our attention. In the discussion section, we focused on identifying some of the limitations related to underpowered studies (personnel needed, amount of time it takes, etc.) – however, power is an important topic to note given the current trends in psychological literature on the science of science. We wish this section to be a guide for researchers who are in this topic area – as you note, an easy to implement design can provide small but important changes for an individual. Therefore, we note that power is a key factor that could really improve in this area to provide more evidence for studies/variables/etc., and we added discussion on how the field can deal with this complex problem (e.g., using crowd sourced research lab opportunities).**

Bottom line: this can be an influential paper. However, it will only be influential if it is read and understood by the readers you seek to influence

\***Thanks much for the insightful and concrete feedback!**

Minor issues. PTS is defined once on page 6 and then never again. It starts bubbling up later in the manuscript. Remind readers what it is periodically,

**\*We have added definitions of PTS in the early methods and discussion sections.**

Reviewer #2: This study conducted a meta-analysis on the pre-post benefits of expressive writing for people diagnosed and not diagnosed with PTSD. Overall, I thought the authors were very methodical in describing their analytic approach and did a good job in summarizing the literature within their introduction. The study did however have several issues that require further attention.

The authors state the general effect size criteria for social science but admit that interpretation of effect sizes should be based on the area of study. If this is the case, then the authors should interpret what is considered a small effect size for expressive writing rather than rely on a general social science standard that fails to take into context the treatment. Why is an effect size of .30 considered small when people are receiving free, easy to do and unintrusive treatment that has virtually no side effects? The authors need to make their justification clear and compelling. Alternatively, they should entertain how a proponent of expressive writing might interpret the magnitude of the effect size estimates they report. I definitely think they should remove the interpretative language from the abstract. Just list the specific effect size values and invite people to read the paper to better understand how you interpret the magnitude of the effect.

**\*One issue with including the effect sizes in the abstract is that there are many ways to estimate effect sizes (as shown in Table 1, 2, 3) depending on theoretical orientation and preferred publication bias adjustor method. We left in the language based on Cohen’s suggestion (as many people use these for better or worse), however, you raise an important point also discussed by the other reviewer. Therefore, we edited the discussion to note that these effects are important in context of the ease of implementation and impact on an individual person’s day to day existence.**

It was also not clear what benefit examining the pre-post treatment effect has over examining the experimental to control comparison. The authors spend 1 sentence describing this component which is arguably one of the most novel contributions of their particular meta-analysis.

**\*This point is discussed for nearly a paragraph in the current meta-analysis section. Additionally, more information was added to the discussion about this choice.**

I was also confused about the particular outcome measures used in the studies they examined. Did ALL of the studies use the exact same instruments to measure post-traumatic stress and growth? If not, then how can we interpret an effect size if a change of 1 point is more meaningful on a 4-point scale than a change of 1 point on a 10 point scale. I am not very familiar with meta-analysis so I hope the authors will provide greater clarification on the instruments and consistency of the instruments used across the studies they examined.

**\*Thank you for the feedback. All of the studies did not use the exact same instruments to measure posttraumatic stress and growth. In the paper, we clarified that we also coded the specific measure that each study employed. In regards to interpretation of effect size, the nature of the Cohen’s d equation allows us to compare across measures. By dividing the mean change by the study standard deviation, Cohen’s d becomes a standardized score for comparison across studies. We have clarified this in the manuscript, as well.**