**Peer Review and Communication History**

**MS Title**: How does Heartbeat Counting Task Performance Relate to Theoretically-Relevant Mental Health Outcomes? A Meta-Analysis

**Author Names**: Olivier Desmedt, Maaike Van Den Houte, Marta Walentynowicz, Sarah Dekeyser, Olivier Luminet, and Olivier Corneille

**Submitted:** Sep 23, 2021

**Editor First Decision**: Revise & Resubmit

Jan 13, 2022

Dear Olivier Desmedt,

First, please accept my apologies for the length of this review process; it was difficult to first find reviewers, and then to get their reviews. At this point, I have received two excellent reviews on your manuscript, and I’m grateful for your patience, and for the thought the reviewers have put into their reviews.

I have now received all reviews of your manuscript, “How does Heartbeat Counting Task Performance Relate to Theoretically-Relevant Mental Health Outcomes? A Meta-Analysis” from qualified researchers. I also independently read the manuscript before consulting these reviews. I agree that your manuscript has important strengths and also that there are some issues that need to be addressed. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

The reviewers did an outstanding job in their reviews. I will highlight issues I think are particularly salient here. In your resubmission, please include a document with a point-by-point response to both the points I list here and the reviewers’ comments, outlining each change made in your manuscript or providing a suitable rebuttal.

In summary, I think this is a promising manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision.

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 contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Kevin King

**Reviewer 1**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

This is an excellent, important, and timely paper. I read it as a preprint some months ago, and enjoyed it immensely.

The authors conduct a rigorous and comprehensive meta-analysis of associations between the most widely used measure of interoception, the heartbeat counting task, and common self-report inventories of anxiety, depression, and alexythymia. Associations with other common moderator variables such as BMI, heart-rate, and age are investigated. In general the authors find essentially no evidence of any association between HCT scores and mental health variables. They do find highly signifigant. but modest effect sizes inter-relating heart-rate and BMI to HCT scores. Finally, p-curve analysis indicates little evidence of p-hacking in the HCT literature.

Overall this is an excellent manuscript and I have very little to add. The logic of the meta-analysis appears to be sound, the study is follows all appropriate methodological guidelines, and the conclusions are highly appropriate given the data. The authors have also followed best practices for transparency in science, including a full pre-registration of the meta-analysis together with open data and code sharing. I particularly like that the authors invite the interoception research community to interrogate the data themselves, as they are (hopefully) sure to cause some concern and controversy!

I have one or two points for consideration in the discussion, and a few other small corrections/suggestions:

1. As I understand it, by design the study examines only estimates of between-subject association (correlation) on HCT and self-reported symptom inventories. While this makes ample sense, I noted that most of the studies seem strongly underpowered (following ref# 1, which suggests such correlation estimates stabilize around 150-250 N for small to moderate effect sizes) to assess this. If we assume that these studies will have very poor estimation of the true correlation size and sign, what does this imply for the meta-analytic estimates? I am not an expert in the technical details of meta-analysis but does one need to worry about “garbage out, garbage in” here? E.g., can the pooling effects of the random effect meta-analysis somehow overcome if the individual study effect sizes are highly unreliable?
2. Similar to the above - I imagine some proponents of the HBC as a clinical measure might respond that though the measure does not pick out such symptoms in healthy subjects, it remains sensitive to group differences between, e.g., patients with major depression disorder or severe clinical anxiety. I think you would certainly expect the self-report inventories to pick this up also given their assumed construct validity, but if the authors do already have data on this, it might be worth including as an exploratory analysis.

Small things:

1. “This is discussed in the remainder of the discussion.” - stylistic thing, maybe “this is considered in the remainder of the discussion” instead.
2. On page 23-24, you could optionally add that Legrand et al (2020) found that HBC under-counting was related to a psychophysical measure of percieved heart-rate, with a more negative bias relating to lower HBC accuracy and higher over-counting. This further supports that HBC is primarily related to a negative bias in heart-rate perception rather than objective sensitivity to heartbeats.

References:

1. Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize?. Journal of Research in Personality, 47(5), 609-612.
2. Legrand, N., Nikolova, N., Correa, C., Brændholt, M., Stuckert, A., Kildahl, N., … & Allen, M. (2021). The heart rate discrimination task: a psychophysical method to estimate the accuracy and precision of interoceptive beliefs. bioRxiv.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  |  | ✔ |

**Reviewer 2**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

This is a timely paper – kudos to the authors. The importance is abundantly clear when looking at Figure 1 (the number of papers using this task per year which has increased exponentially). Clearly a great deal of work went into this paper and I am happy have been invited to review it. I have no major concerns, only suggested improvements.
Introduction:

1. Whilst the introduction is nice and concise, I think that it is worth acknowledging some of the controversy surrounding the HCT to further justify this meta-analysis. This is mentioned in the discussion, but I think it is important rationale to include here (briefly).
2. The authors also state “these associations are strongly expected on a theoretical level”. This is true, if we accept that the HCT measures interoception. I think by addressing the previous suggestion the authors could make it more explicit that here we are not assessing whether interoception relates to these factors, but whether this task specifically does (or not). Whilst those who know the field will be aware of this, for a reader not so familiar with this task this would not be clear.
Method:
3. I was a little surprised that aggregate scores were removed (e.g., healthy vs. typical). I am not sure why you would expect different effects (e.g., opposite effects) in each population. Is there evidence to suggest this? Doesn’t splitting these groups reduce the size of the distribution of the other measures (e.g., anxiety etc.,) which may reduce the possibility of observing an effect?
4. Did they authors extract any information about the HCT administration? E.g., instructions used, number of trials, analysis (Hart vs Schandry) scoring methods? It seems this may have been done based on later sections, but this is not reported here. Adding this would help clarity.
5. Did the authors attempt to contact authors for unreported data? If not, was there a reason why this was not attempted? This could be added to the paper.
Results:
6. In the methods the authors say they did influence analysis (e.g., to assess the reliability of the effect size). Was this the leave one out method? If not, that is often very useful to include to see how stable the effect size is.
Discussion:
7. “They represented very small effect sizes that may be explained by non-interoceptive processes.” I don’t disagree that these are very small, but the meta analysis does not tell us whether effects are driven by non interoceptive processes. This should be removed from the first discussion paragraph and can be speculated on later.
8. As in the introduction, I think it is important to be very careful to be clear that conclusions are regarding relationships using this measure. I think if the introduction sets the scene for this (e.g., problems with the HCT), there will not be a risk of assuming this applies to interoception more broadly. It may do of course and we’ll all be looking for a career change, but given the small relationships across different tasks (as the authors note) it is important to caveat this.
9. More of a theoretical point – even if cardiac signal intensity underlies differences, this would be an interoceptive process not a confound. This is also mentioned in the limitations, but again I don’t think it is a confound. Individual differences in interoception may be due to the perception of the signal, which may depend upon ability to ‘tap into’ the signal as well as the size of the signal. I think the key question regarding age/sex is whether a domain general or specifically interoceptive deficit underlies differences. I think that the findings of sex differences and age require a more careful interpretation. These needn’t suggest validity of the task per se, but at present the explanation for these is weaker in comparison to the other areas.
10. It might be worth drawing out the conclusions about the HCT a little more. Do we find these relationships using other tasks? Is the HCT simply invalid because it has been administered unreliably? I think that to be fully convincing it would be useful for the authors to address the elephant in the room – is interoception the problem (e.g., theory is wrong, accuracy doesn’t matter – but perhaps other aspects of interoception do), is the measure the problem (the HCT should be abandoned) or is the measure OK if administered with strict instructions. Clearer sectioning (e.g., considering of each possibility explicitly) would help to arrive at a clearer conclusion (or at least outline the possible conclusions).
11. The authors suggest that the STAI might be a problem – could the authors not look at the association as assessed by other measures to check this using their data? My hunch is that this might make little difference (which would allow that to be ruled out).

Minor:
Abstract

1. It is worth noting that the HCT measures cardiac interoception in the abstract. As interoception dissociates across channels, this would be helpful clarity.
2. 11.524 participants – should this be a comma?

Method:
3. The authors might wish to consider using the more recent PRISMA checklist (from 2020).
4. “We included studies with adult participants (18+) and investigating the association” (page 4) should this be ‘that’ investigated.
Discussion:
5. e.g., vision, hearing, proprioception, and pain perception – notably pain perception is interoceptive under some models.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

**Author Response**
Feb 21, 2022

Dear Dr. King,

It is our pleasure to submit to *Collabra* the revision of our manuscript entitled “How does Heartbeat Counting Task Performance Relate to Theoretically-Relevant Mental Health Outcomes? A Meta-Analysis”. We would like to thank you for giving us the opportunity to improve our manuscript by addressing the comments and suggestions provided in your editorial correspondence. We addressed all reviewers’ comments. We briefly summarize here the main revisions implemented in this revised version: (1) we now specify in the introduction that the HCT is characterized by validity issues, (2) we further discuss that our findings could either suggest that current theoretical assumptions are incorrect (i.e., no relationship exists between IAcc and these mental health variables), that the HCT scores do not adequately capture individual differences in cardiac IAcc, or both, (3) the discussion around the relationship between HCT scores and age/sex has been substantially expanded, and (4) other minor revisions were made.

We would like to thank you and the reviewers for the many constructive comments indicated in the editorial correspondence. We hope that this revision will meet with your approval. If not, we would be happy to address any lingering concern or clarification request.

Yours sincerely,

Olivier Desmedt, Maaike Van Den Houte, Marta Walentynowicz, Sarah Dekeyser, Olivier Luminet, & Olivier Corneille

**Reviewer # 1**

*This is an excellent, important, and timely paper. I read it as a preprint some months ago, and enjoyed it immensely.*

*The authors conduct a rigorous and comprehensive meta-analysis of associations between the most widely used measure of interoception, the heartbeat counting task, and common self-report inventories of anxiety, depression, and alexithymia. Associations with other common moderator variables such as BMI, heart-rate, and age are investigated. In general, the authors find essentially no evidence of any association between HCT scores and mental health variables. They do find highly significant. but modest effect sizes inter-relating heart-rate and BMI to HCT scores. Finally, p-curve analysis indicates little evidence of p-hacking in the HCT literature.*

*Overall this is an excellent manuscript and I have very little to add. The logic of the meta-analysis appears to be sound, the study is follows all appropriate methodological guidelines, and the conclusions are highly appropriate given the data. The authors have also followed best practices for transparency in science, including a full pre-registration of the meta-analysis together with open data and code sharing. I particularly like that the authors invite the interoception research community to interrogate the data themselves, as they are (hopefully) sure to cause some concern and controversy!*

**Response:** We would like to thank the reviewer for his/her very positive feedback on our manuscript.

1. *I have one or two points for consideration in the discussion, and a few other small corrections/suggestions: As I understand it, by design the study examines only estimates of between-subject association (correlation) on HCT and self-reported symptom inventories. While this makes ample sense, I noted that most of the studies seem strongly underpowered (following ref# 1, which suggests such correlation estimates stabilize around 150-250 N for small to moderate effect sizes) to assess this. If we assume that these studies will have very poor estimation of the true correlation size and sign, what does this imply for the meta-analytic estimates? I am not an expert in the technical details of meta-analysis but does one need to worry about “garbage out, garbage in” here? E.g., can the pooling effects of the random effect meta-analysis somehow overcome if the individual study effect sizes are highly unreliable?.*

**Response:** This is a very relevant question. Meta-analyses composed of small sample size studies can indeed lead to estimation bias. To address this issue, we implemented two recommended solutions: First, we transformed correlation coefficients into Fisher’s z, which allows removing the range restriction and making the sampling distribution approximately normal. This is because correlations are restricted in their range, which can lead to bias when estimating the standard error for studies with a small sample size (Alexander et al., 1989). Second, we gave more weight to studies with higher sample size and lower standard error, as recommend by Hedge and Olkin (2014). Despite the use of these methods, bias can still remain (Lin, 2018). However, this is more likely to be true if effect size is correlated with sample size, which would represent a publication bias. Reassuringly, our analyses showed that most associations were not affected by publication bias.

1. *Similar to the above - I imagine some proponents of the HBC as a clinical measure might respond that though the measure does not pick out such symptoms in healthy subjects, it remains sensitive to group differences between, e.g., patients with major depression disorder or severe clinical anxiety. I think you would certainly expect the self-report inventories to pick this up also given their assumed construct validity, but if the authors do already have data on this, it might be worth including as an exploratory analysis.*

**Response:** Indeed, the HCT could be predictive of individual differences in mental health outcomes among clinical populations only. This is the reason why we conducted moderation analyses with the type of population (healthy vs. clinical) as a moderator; this moderation was non-significant, which does not support this hypothesis.

1. *Small things:*
	1. *“This is discussed in the remainder of the discussion.” - stylistic thing, maybe “this is considered in the remainder of the discussion” instead.*
	2. *On page 23-24, you could optionally add that Legrand et al (2020) found that HBC under-counting was related to a psychophysical measure of percieved heart-rate, with a more negative bias relating to lower HBC accuracy and higher over-counting. This further supports that HBC is primarily related to a negative bias in heart-rate perception rather than objective sensitivity to heartbeats.*

**Response:** This has been amended (p.20 and p.25).

**Reviewer #2**

*This is a timely paper – kudos to the authors. The importance is abundantly clear when looking at Figure 1 (the number of papers using this task per year which has increased exponentially). Clearly a great deal of work went into this paper and I am happy have been invited to review it. I have no major concerns, only suggested improvements.*

**Response:** We would like to thank the reviewer for his/her positive feedback on our manuscript.

1. *Introduction: Whilst the introduction is nice and concise, I think that it is worth acknowledging some of the controversy surrounding the HCT to further justify this meta-analysis. This is mentioned in the discussion, but I think it is important rationale to include here (briefly).*
2. *The authors also state “these associations are strongly expected on a theoretical level”. This is true, if we accept that the HCT measures interoception. I think by addressing the previous suggestion the authors could make it more explicit that here we are not assessing whether interoception relates to these factors, but whether this task specifically does (or not). Whilst those who know the field will be aware of this, for a reader not so familiar with this task this would not be clear.*

**Response:** We agree. These two pieces of information have now been added to the introduction (p.2-3 and 3-4).

1. *Method: I was a little surprised that aggregate scores were removed (e.g., healthy vs. typical). I am not sure why you would expect different effects (e.g., opposite effects) in each population. Is there evidence to suggest this? Doesn’t splitting these groups reduce the size of the distribution of the other measures (e.g., anxiety etc.,) which may reduce the possibility of observing an effect?*

**Response:** We decided from the very beginning to exclude correlation coefficients computed on aggregated populations (i.e., healthy and clinical populations) because we were interested in the moderation effect of population type. Previous studies have indeed suggested that psychological processes and symptoms show differential connectivity in clinical (vs. healthy) populations (Fried et al., 2017). We agree that restriction of range can bias (towards an under or over-estimation) correlation coefficients. However, we do not see how this issue could have applied to our current meta-analysis. When performing our moderation analysis, we included data coming from both populations. Therefore, this range restriction does not apply. In addition, range restriction should have an effect only if slopes (or correlation coefficients) differ across ranges (here, the “healthy” and the “clinical” ones). Our moderation analysis indicates no significant effect of population type.

1. *Did they authors extract any information about the HCT administration? E.g., instructions used, number of trials, analysis (Hart vs Schandry) scoring methods? It seems this may have been done based on later sections, but this is not reported here. Adding this would help clarity.*

**Response:** We indeed extracted all procedural details related to HCT administration, including the instructions, number of trials and time intervals, heart rate device, statistical formula, and the presence of a training phase. This information is available in open access (https://osf.io/3cwt4/?view\_only=7e466bd994134dc787cc6a778f1d0723). This has now been specified in the manuscript (p.6).

1. *Did the authors attempt to contact authors for unreported data? If not, was there a reason why this was not attempted? This could be added to the paper.*

**Response:** We did not contact the authors for unreported data because past experiences taught us that researchers often are not willing to share those, which introduces delays and may generate frustration on both sides. This information has been added to the manuscript (p.8). However, because no publication bias was evidenced on the study set included for this meta-analysis, we are relatively confident that our conclusions should not reflect a file drawer effect (here, people keeping significant results in their drawers - as few meta-analytic effects were obtained overall).

1. *Results: In the methods the authors say they did influence analysis (e.g., to assess the reliability of the effect size). Was this the leave one out method? If not, that is often very useful to include to see how stable the effect size is..*

**Response:** We indeed performed the leave-one-out method. This information was reported in the manuscript (p.7).

1. *Discussion: “They represented very small effect sizes that may be explained by non-interoceptive processes.” I don’t disagree that these are very small, but the meta-analysis does not tell us whether effects are driven by non interoceptive processes. This should be removed from the first discussion paragraph and can be speculated on later.*

**Response:** We agree, the end of the sentence has been removed from the manuscript.

1. *As in the introduction, I think it is important to be very careful to be clear that conclusions are regarding relationships using this measure. I think if the introduction sets the scene for this (e.g., problems with the HCT), there will not be a risk of assuming this applies to interoception more broadly. It may do of course and we’ll all be looking for a career change, but given the small relationships across different tasks (as the authors note) it is important to caveat this.*

**Response:** We agree. This has been clarified in the introduction and discussion (p.3-4 and 21-22).

1. *More of a theoretical point – even if cardiac signal intensity underlies differences, this would be an interoceptive process not a confound. This is also mentioned in the limitations, but again I don’t think it is a confound. Individual differences in interoception may be due to the perception of the signal, which may depend upon ability to ‘tap into’ the signal as well as the size of the signal. I think the key question regarding age/sex is whether a domain general or specifically interoceptive deficit underlies differences. I think that the findings of sex differences and age require a more careful interpretation. These needn’t suggest validity of the task per se, but at present the explanation for these is weaker in comparison to the other areas.*

**Response:** We respectfully disagree. For us, as cardiac interoceptive accuracy is defined as the objective capacity to detect cardiac signals, and the HCT is meant to be a measure of this capacity, it should not be contaminated by the intensity of the signal. As a comparison, let’s imagine that “an ophthalmologist uses a Chart Test with letters for measuring visual accuracy among patients. The eyes-to-chart distance, however, varies across patients (the signal is stronger for some patients than others). The doctor concludes that visual acuity is better for patients sitting two meters closer to the chart. We believe many would question the ophthalmologist’s assessment” (Corneille et al., 2020). Consistently, we think it is questionable to conflate the capacity to detect heartbeats with their intensity. Nevertheless, we now acknowledge in the discussion that other researchers may think otherwise (p.22).

The discussion on the relationship between HCT scores and age/sex has been substantially expanded (p.22-23).

1. *It might be worth drawing out the conclusions about the HCT a little more. Do we find these relationships using other tasks? Is the HCT simply invalid because it has been administered unreliably? I think that to be fully convincing it would be useful for the authors to address the elephant in the room – is interoception the problem (e.g., theory is wrong, accuracy doesn’t matter – but perhaps other aspects of interoception do), is the measure the problem (the HCT should be abandoned) or is the measure OK if administered with strict instructions. Clearer sectioning (e.g., considering of each possibility explicitly) would help to arrive at a clearer conclusion (or at least outline the possible conclusions).*

We agree with the reviewer that our findings could suggest that the HCT lacks validity, or that the theories should be qualified, or both (besides other methodological and statistical limitations that would concern most studies); this was mentioned in our conclusion. Nevertheless, we think that our data do not allow us to answer this question and attempting to do so would be highly speculative. The original version already included a discussion about existing theoretical assumptions and HCT limitations. Furthermore, we do not think that modifying the HCT administration (e.g., the instructions) would completely resolve this validity issue as another important bias (i.e., response bias) still applies, as already explained in p.26.

However, in accordance with the reviewer’s suggestion, we now further discuss the different plausible interpretations (p.21-22). We also encourage future studies to use alternative measures of IAcc to answer this question (given that very few previous studies did so; e.g., Plans et al., 2021; Van Den Houte et al., 2021)

1. *The authors suggest that the STAI might be a problem – could the authors not look at the association as assessed by other measures to check this using their data? My hunch is that this might make little difference (which would allow that to be ruled out).*

Indeed, moderation analyses showed that our conclusions did not change depending on the questionnaire used (i.e., STAI vs other anxiety questionnaires but also BDI-II vs. other depression questionnaires). This information has been added to the manuscript (p.28).

1. *Minor:*

*Abstract:*

*It is worth noting that the HCT measures cardiac interoception in the abstract. As interoception dissociates across channels, this would be helpful clarity.*

*11.524 participants – should this be a comma?*

*Method:*

*The authors might wish to consider using the more recent PRISMA checklist (from 2020).*

*“We included studies with adult participants (18+) and investigating the association” (page 4) should this be ‘that’ investigated.*

*Discussion:*

*e.g., vision, hearing, proprioception, and pain perception – notably pain perception is interoceptive under some models.*

**Response:** This has been amended in the manuscript (p.1, p.4, & p.23). The PRISMA checklist has been updated. We thank the reviewer for her/his thorough review.

**References:**

Alexander, R. A., Scozzaro, M. J., & Borodkin, L. J. (1989). Statistical and empirical examination of the chi-square test for homogeneity of correlations in meta-analysis. *Psychological Bulletin*, *106*(2), 329–331. https://doi.org/10.1037/0033-2909.106.2.329

Corneille, O., Desmedt, O., Zamariola, G., Luminet, O., & Maurage, P. (2020). A heartfelt response to Zimprich et al. (2020), and Ainley et al. (2020)’s commentaries: Acknowledging issues with the HCT would benefit interoception research. *Biological Psychology*, *152*, 107869. https://doi.org/10.1016/j.biopsycho.2020.107869

Fried, E. I., van Borkulo, C. D., Cramer, A. O. J., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: A review of recent insights. *Social Psychiatry and Psychiatric Epidemiology*, *52*(1), 1–10. https://doi.org/10.1007/s00127-016-1319-z

Hedges, L. V., & Olkin, I. (2014). *Statistical Methods for Meta-Analysis*. Academic Press.

Lin, L. (2018). Bias caused by sampling error in meta-analysis with small sample sizes. *PLOS ONE*, *13*(9), e0204056. https://doi.org/10.1371/journal.pone.0204056

Plans, D., Ponzo, S., Morelli, D., Cairo, M., Ring, C., Keating, C. T., Cunningham, A. C., Catmur, C., Murphy, J., & Bird, G. (2021). Measuring interoception: The phase adjustment task. *Biological Psychology*, *165*, 108171. https://doi.org/10.1016/j.biopsycho.2021.108171

Van Den Houte, M., Vlemincx, E., Franssen, M., Diest, I. V., Oudenhove, L. V., & Luminet, O. (2021). The respiratory occlusion discrimination task: A new paradigm to measure respiratory interoceptive accuracy. *Psychophysiology*, *58*(4), e13760. https://doi.org/10.1111/psyp.13760

**Editor Final Decision:** Accept

Feb 28, 2022

Dear Olivier Desmedt,

I have now had a chance to read over your manuscript “How does Heartbeat Counting Task Performance Relate to Theoretically-Relevant Mental Health Outcomes? A Meta-Analysis”, along with the letter describing the changes you made. Thank you for your responsiveness to the concerns that the reviewers and I raised. I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point.

You will be receiving separate correspondence regarding any production and technical comments, data deposits, as well as publication charges. We work with the Copyright Clearance Center to process any applicable APC charges. Please note that your APC transaction must be completed before your article gets published.

You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,
Kevin King