**Peer Review and Communication History**

**MS Title**: The Convergence of Self and Informant Reports in a Large Online Sample

**Author Names**: Anne Zola, David M. Condon, William Revelle

**Submitted**: 2020

**Editor First Decision—Revise & Resubmit**

May 8, 2020

Dear Ms Anne Zola,

Thank you for submitting your work to Collabra: Psychology. Foremost, I apologize for the delay in processing this paper. This was a busy semester and then the pandemic hit. I was only able to secure reviewer from one source. However, the reviewer is an outstanding scholar who is quite knowledgeable about personality trait structure, online personality measurement, and self-other issues. I thank them for their service to this journal. I independently read the paper and then consulted the comments. Given that I have some knowledge about the topic of the paper and the high quality of the review, I believed I could move forward with a decision rather than draw out the process.

As you will read below, the reviewer was generally positive above the work. I was also positive about the work and I think there is the raw material to make an interesting contribution to the literature. Thus, I will extend a revise and resubmit decision. I do not plan to send the revision to back to the reviewer so the time to make a final decision should be reduced if you indeed elect to submit a revision. I will say that there are several issues that warrant attention. It might take some work to craft a suitable revision, but I think it is achievable.

The single reviewer raised excellent points and I their views are representative of the reactions of other readers. You should address each of their concerns either in the revised text or in the response letter. Addressing those concerns will make this a stronger paper. I will comment on the most salient points and offer some of my own reactions when reading the paper in my letter. You might disagree with some (or all) of the points so feel free to pushback against any suggestions you believe will harm your work. Please just describe your counterpoints in the return letter.

1. The paper is short, and I think could be expanded to provide readers greater context for the ideas and especially the underlying structural model for the personality scales. Put differently, I think more background about the SPI model is needed so readers understand the structural model that serves as the basis for the work.

2. The reviewer points out how the two stated goals of the paper are not necessarily fully achieved. I think this means more discussion and likely ancillary analyses. So please take that reaction seriously.

3. In terms of feasibility, the rate of data collection is slow and requires informants to opt-in. The approach described in the paper could be used but perhaps the data yield is not as great as with self-reports and feedback. This point could be acknowledged and some directions for the future could be offered. Would incentives help? Providing more or different kinds of feedback to informants? How do you think the process could be improved or scaled up or extended to say research with other populations?

4. As the reviewer noted, the different item sets can make some of the self-other correlations harder to interpret. Please attend to the suggestions offered by the reviewer and thoroughly discuss the pros/cons with the approach used here.

5. I think the raw self-other correlations should be compared to the ones corrected or adjusted for reliability (I think I sort of dislike the term corrected but I know that is industry standard). Regardless, that process needs more description. What reliability estimates were used? Why? Should alpha or omega be used or even short-term dependability data? To be clear, I don’t necessarily have an issue with adjusting the coefficients, but the raw data should be featured along-side the adjusted ones and the issues with corrections/adjustments for attenuation should be discussed. The reviewer makes this point as well.

6. I think Table 1 needs a note to explain the negative correlations (reverse-coded items). It took me a second to register this as my first reaction was being puzzled when I saw negative signs. (I guess this could be my cognitive limitations, but a simple table note could help). I think the heading for the r column should read Item-Scale Correlation.

7. Should heterarchical be hierarchical on page 8?

8. I think adding headings to indicate Self and Informant sources would make some tables easier to read. So, for example, can you add an overarching column heading to Table 4 that says Informant Reports to cover columns 2 to 9 and Self-Report to cover the first column? I think additions like this will make the tables easier to apprehend.

9. I think the numbers in parentheses in Table 6 need to be explained in the table note.

10. It might be useful to provide a short section giving your framework for interpreting effect sizes (e.g., 10 is small based on X, Y, and Z).

Those were the big issues based on my reading. I also noted many of the same issues as the Reviewer that I did not cover in #1 to #10 so please attend to all of their points if you revise this work. And I do hope you revise this paper!

To access your submission account, follow the below instructions:
1) login to the journal webpage with username and password
2) click on the submission title
3) click 'Review' menu option
4) download Reviewed file and make revisions based on review feedback
5) upload the edited file
6) Click the 'notify editor' icon and email the confirmation of re-submission and any relevant comments to the journal.

If you have any questions or difficulties during the revision process, please contact us. Thanks again for considering Collabra as a showcase for your interesting work. I hope you are all doing well given the conditions in the world.

Sincerely,

Brent Donnellan
Michigan State University
donnel59@msu.edu

------------------------------------------------------
Reviewer B:

1) General comments and summary of recommendation
Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:
 - Are the methodologies used appropriate?
 - Are any methodological weaknesses addressed?
 - Is all statistical analysis sound?
 - Does the conclusion (if present) reflect the argument, is it supported by data/facts?
 - Is the article logically structured, succinct, and does the argument flow coherently?
 - Are the references adequate and appropriate?:
This paper examines associations between personality self-reports and informant-reports collected online. The paper’s strengths include its large sample and use of the SAPA website to collect data on an unusually large and diverse set of test items. I therefore think it has the potential to make both methodological and substantive contributions to the self-peer agreement literature. I also have some concerns about the paper, but it may be possible to address all of these in a thoughtful revision.

Main Points

1. The paper emphasizes two possible contributions. The first is showing that it is feasible to collect large samples of personality informant-reports online. The authors have indeed collected a large sample of informant-reports through the (very cool) SAPA website. However, I was struck by the fact that fewer than 1% of potential participants who completed a SAPA self-report opted into the informant-report portion of the study. As a result, the sample of 921 unique targets was accumulated at the rate of one every couple of days. This would be a pretty slow timeline for many research projects. I therefore recommend that the authors (a) acknowledge the rather slow pace of accumulating informant-reports through the opt-in method used in this study, and (b) offer any advice they may have for accelerating this pace in future research.

2. The second possible contribution is advancing our understanding of self-peer and peer-peer agreement in personality reports. My main concern here is that the self-reports and peer-reports were based on different items, and in some cases different constructs. This makes the self-peer coefficients difficult to interpret, because they may be suppressed by both (a) genuine disagreement between the self and peers, and (b) methodological differences between the self and peer measures. I recommend acknowledging and discussing this issue. It may also be possible to address the issue directly: If the self-report item pool includes the informant-report items, then how are the results affected by limiting the analyses to the overlapping self and peer items?

3. The internal consistency reliabilities of some peer-report scales are rather low (p. 7), which means that correcting the agreement coefficients for reliability (p. 16) could lead to quite large adjustments. This left me wondering (a) whether the low internal consistencies were intentional or surprising, and (b) how much correcting for reliability affected the self-peer agreement coefficients. I therefore recommend discussing (a) and reporting (b).

4. The Results section is light on interpretation. This may be intentional (i.e., letting the results speak for themselves), but I think it would be helpful for readers to add a couple sentences at the end of each Results subsection (e.g., Group differences) to summarize and highlight the key findings from those analyses.

5. I recommend adding a Discussion subsection about limitations and future directions. This subsection could help address the points raised above, as well as any other limitations and future directions that the authors would like to consider.

Additional Points

p. 2 and throughout. In a few places (starting with the Abstract), the authors state that they examined self-peer agreement at the “item level.” This led me to expect agreement coefficients between individual self-report items and peer-report items. However, the analyses estimate correspondence between self-report items and peer-report scales. I recommend clarifying this early in the manuscript, or possibly choosing a different term than item-level agreement.

p. 9. The SAPA’s massively missing design is very cool, and the total sample size of 158k self-reports is impressive. However, it left me wondering: What was the obtained sample size for pairwise item correlations? I recommend clarifying.

2) Figures/tables/data availability:
Please comment on the author’s use of tables, charts, figures, ifrelevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here).:
I recommend adding a caption to the figure. Other than that, the tables are relevant and the data are available.

3) Ethical approval:
If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared.
If not, please detail where you think a further ethics approval/statement/follow-up is required.:
I did not notice a statement about ethics approval. I recommend adding one.

4) Language:
Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.:
The text is generally well written, but I recommend expanding and clarifying a few points about the methodology and results:

“It should be noted that this self-report assessment model was not supported by the evidence” (p. 8).

“This matrix was scored adjusting for overlapping keys and replacing the overlapping covariances with the corresponding best estimate of the item’s “true variance,” the average correlation for that item” (p. 9).

“Items were included as “best items" if they emerged for all ten iterations of k-fold cross validation” (p. 9).

p. 18. “Moreover, the structure of this self-informant agreement corroborates the overlapping structure of the 27 lower-level factors in the SPI.” How so?

Also, I think the Method section would be easier to follow if the Participants subsection were placed before the Analysis subsection.

**Author Response**

July 9, 2021

Dear Dr. Donnellan,

Thank you for the opportunity to revise the attached manuscript, *The Convergence of Self and Informant Reports in a Large Online Sample.* We appreciate the helpful suggestions from yourself and the reviewer. This document details the steps we have taken in response to each suggestion or request. Editor/reviewer comments are in italics and our responses are included below each.

**Editor Comments**

*1.      The paper is short, and I think could be expanded to provide readers greater context for the ideas and especially the underlying structural model for the personality scales.  Put differently, I think more background about the SPI model is needed so readers understand the structural model that serves as the basis for the work.*

 This is a good suggestion, and we agree that the prior version was lacking sufficient explanation of the SPI measure. We were originally concerned that such discussion might be distracting in the context of the methods used to collect the informant-report data, but we have put considerable effort into re-drafting the Method section (and the Measures sub-section, in particular) to add this content in a clear manner. We hope you will agree that this request has been achieved, especially in second and third paragraphs of the Measures sub-section on pages 8-9 of the current version.

*2.      The reviewer points out how the two stated goals of the paper are not necessarily fully achieved.  I think this means more discussion and likely ancillary analyses.  So please take that reaction seriously.*

 These two goals mentioned in the manuscript relate to demonstrating the feasibility of collecting large samples of informant reports online and providing evidence for the validity of the SPI with self-informant agreement. Regarding the first point, we have added additional discussion of the limitations of the data collection method including metrics to describe the rate of data return for informant- and self-report. We also now provide multiple recommendations for future research using an online informant-ratings. To underscore the second point, we have elaborated further on the comparative validity of the self-informant associations for the SPI-5 relative to those of the BFAS, the IPIP Big Five, and the IPIP-NEO. Specifically, we have added text associated with these analyses in the Method and Results sections. In the discussion, we added the following text (on p. X of the current version):

When comparing self-informant agreement of the SPI to the other operationalizations of the Big Five (BFAS, IPIPB5, and IPIPNEO), the findings are more similar than different. The most prominent differences are for Openness/Intellect, where the SPI and IPIPNEO self-reports are both less highly associated with the informant reports than the BFAS and IPIPB5. These pairings are unsurprising given the derivation of the BFAS from the IPIP Big Five (DeYoung et al., 2007), but the lower magnitudes – especially for SPI and NEO Openness – reflect the reality that longer self-report measures of this trait in particular include more diverse content than is covered by the 4-item informant-report. The SPI Extraversion self-informant correlation is also somewhat lower than the other models (.63 vs .70-.76), but this can once again be explained by the lack of empirical evidence supporting the inclusion of Assertiveness in the derivation of the SPI (Condon, 2018). In fact, this is an important theme of the findings reported here as there was considerable content included in the pools of items used to derive the SPI, including all of the content from the BFAS, IPIPB5, and the IPIPNEO (Condon, 2018).

Of note, the reviewer suggested that the second goal of the paper was to advance our understanding of self-peer and peer-peer agreement. Though we did not explicitly state this, we acknowledge that this is definitely an important goal. The reviewer raised a couple of other important points – these are addressed directly below.

*3.      In terms of feasibility, the rate of data collection is slow and requires informants to opt-in.  The approach described in the paper could be used but perhaps the data yield is not as great as with self-reports and feedback.  This point could be acknowledged and some directions for the future could be offered.  Would incentives help?  Providing more or different kinds of feedback to informants?  How do you think the process could be improved or scaled up or extended to say research with other populations?*

This is also a very good suggestion. We have added considerable discussion of the limitations of the data collection method and recommendations for future research that use similar processes. See pages 24 and 25 of the current version.
 *4.      As the reviewer noted, the different item sets can make some of the self-other correlations harder to interpret.  Please attend to the suggestions offered by the reviewer and thoroughly discuss the pros/cons with the approach used here.*

We agree that the different self and informant items increases the difficulty of interpreting results and have now included additional discussion of this limitation in multiple places in the manuscript, specifically the Method and throughout our interpretation of the analyses in the Results and Discussion. We also added the following text to the Limitations sub-section at the end of the paper:

In this study, we discuss the convergence of self and informant report items, but it should be reiterated the self- and informant-report assessment models were dissimilar by design. As discussed above, the informant survey was based on a theory-driven “aspects” model of the Big Five/Big Six domains whereas the self-report instrument was derived empirically based on the structure underlying a large pool of personality items. This is a limitation because it led to an imperfect evaluation of self- and informant-ratings based on common structure. This means that the results were conflated beyond the simple self/other distinction because we are also mapping across different frameworks. Future research should consider a more direct evaluation of the relations between self- and informant-ratings using an informant-report model that matches the self-report SPI model more directly. Note that using 2-item observer report scales for 27 factors would more than double the burden on observers relative to the 24 items administered here.

*5.      I think the raw self-other correlations should be compared to the ones corrected or adjusted for reliability (I think I sort of dislike the term corrected but I know that is industry standard).  Regardless, that process needs more description.   What reliability estimates were used?  Why?  Should alpha or omega be used or even short-term dependability data? To be clear, I don’t necessarily have an issue with adjusting the coefficients, but the raw data should be featured along-side the adjusted ones and the issues with corrections/adjustments for attenuation should be discussed. The reviewer makes this point as well.*

 We have now clarified that corrections were made using standardized alphas and discussed the caveats of interpreting corrected correlations when reliabilities of some informant scales were relatively low. We also added a table to show the unadjusted correlations alongside the corrected correlations, as requested.

*6.      I think Table 1 needs a note to explain the negative correlations (reverse-coded items).  It took me a second to register this as my first reaction was being puzzled when I saw negative signs.  (I guess this could be my cognitive limitations, but a simple table note could help).  I think the heading for the r column should read Item-Scale Correlation.*

Thank you for these suggestions. We have added a note to Table 1 note clarifying the negative correlations are present for reverse-coded items. We have also changed the column heading to “Item-scale correlation” in this Table and in Table 6.

*7.      Should heterarchical be hierarchical on page 8?*

We have changed heterarchical to hierarchical on page 8, as we agree it more clearly describes the structure. On page 14., we mention the term heterarchical again, this time with a reference to the Milyavaskaya et al. (2013) from whom we borrow the term.

Milyavskaya, M., Philippe, F. L., & Koestner, R. (2013). Psychological need satisfaction across levels of experience: Their organization and contribution to general well-being. *Journal of Research in Personality, 47*(1), 41-51.

*8.      I think adding headings to indicate Self and Informant sources would make some tables easier to read.  So, for example, can you add an overarching column heading to Table 4 that says Informant Reports to cover columns 2 to 9 and Self-Report to cover the first column?  I think additions like this will make the tables easier to apprehend.*

We have added overarching column headings to Tables 4, 5, and 6 to improve the readability.

*9.      I think the numbers in parentheses in Table 6 need to be explained in the table note.*

We have added a note in this Table (now Table 8) that mean item-scale correlations are presented in the paratheses.  *10.     It might be useful to provide a short section giving your framework for interpreting effect sizes (e.g., 10 is small based on X, Y, and Z).*

We have added language in the Results sections describing how we interpret effect sizes based both on Cohen’s framework and empirical evidence in this area. In both domains, a Cohen’s d of 0.20 is considered a small effect, and therefore, a 0.10 could be considered a very small effect. We only discuss results where the effect size was 0.10, but emphasize that readers should interpret results with respect of the effect size and 95% confidence intervals.

**Reviewer B**

Main Points

*1. The paper emphasizes two possible contributions. The first is showing that it is feasible to collect large samples of personality informant-reports online. The authors have indeed collected a large sample of informant-reports through the (very cool) SAPA website. However, I was struck by the fact that fewer than 1% of potential participants who completed a SAPA self-report opted into the informant-report portion of the study. As a result, the sample of 921 unique targets was accumulated at the rate of one every couple of days. This would be a pretty slow timeline for many research projects. I therefore recommend that the authors (a) acknowledge the rather slow pace of accumulating informant-reports through the opt-in method used in this study, and (b) offer any advice they may have for accelerating this pace in future research.*

We have added additional discussion of the limitations of this approach and provide recommendations for future research that could accelerate the rate of data collection.

*2. The second possible contribution is advancing our understanding of self-peer and peer-peer agreement in personality reports. My main concern here is that the self-reports and peer-reports were based on different items, and in some cases different constructs. This makes the self-peer coefficients difficult to interpret, because they may be suppressed by both (a) genuine disagreement between the self and peers, and (b) methodological differences between the self and peer measures. I recommend acknowledging and discussing this issue. It may also be possible to address the issue directly: If the self-report item pool includes the informant-report items, then how are the results affected by limiting the analyses to the overlapping self and peer items?*

As stated above we have included additional discussion of the pros and cons of the different self and informant items. We appreciate the suggestion to compare the matching items directly, however informant-report items were not included in the self-report item pool precluding a direct comparison. We apologize if the prior draft created confusion about this point.

 *3. The internal consistency reliabilities of some peer-report scales are rather low (p. 7), which means that correcting the agreement coefficients for reliability (p. 16) could lead to quite large adjustments. This left me wondering (a) whether the low internal consistencies were intentional or surprising, and (b) how much correcting for reliability affected the self-peer agreement coefficients. I therefore recommend discussing (a) and reporting (b).*

 We now discuss how low reliabilities for informant scales are likely to due to the short length of the scale and corrected correlations should be interpreted with the caveat that low reliabilities could lead to large adjustments. We now include raw correlations (corrected for item overlap but not scale reliability) in text and in Table 5 to allow for direct comparison of the corrected and raw correlations.

*4. The Results section is light on interpretation. This may be intentional (i.e., letting the results speak for themselves), but I think it would be helpful for readers to add a couple sentences at the end of each Results subsection (e.g., Group differences) to summarize and highlight the key findings from those analyses.*

The Reviewer was correct to assume that we wished to have the results speak for themselves by keeping the Results section brief. However, we agree syntheses of the key findings would improve the Results section and have now added a couple summarizing sentences to each subsection.

*5. I recommend adding a Discussion subsection about limitations and future directions. This subsection could help address the points raised above, as well as any other limitations and future directions that the authors would like to consider.*

We have added a limitations and future directions section to the discussion where we have addressed many of the points raised by the reviewer and editor.

Additional Points

*p. 2 and throughout. In a few places (starting with the Abstract), the authors state that they examined self-peer agreement at the “item level.” This led me to expect agreement coefficients between individual self-report items and peer-report items. However, the analyses estimate correspondence between self-report items and peer-report scales. I recommend clarifying this early in the manuscript, or possibly choosing a different term than item-level agreement.*

We agree with the Reviewer the term “item-level” was confusing as the analyses we were referring to used self-report items and informant-report scales (not items). To make this clearer, we changed instances of “item-level” to “item-scale” or “item-scale level” where appropriate.

*p. 9. The SAPA’s massively missing design is very cool, and the total sample size of 158k self-reports is impressive. However, it left me wondering: What was the obtained sample size for pairwise item correlations? I recommend clarifying.*

*We have added text addressing the point specifically in the Methods section, including some context for comparing the values to recommendations from prior work. The text reads:*

The number of pairwise administrations between all of the self-report items in these frameworks (minimum = 1,988; *Mdn* = 2,770; *m* = 3,128.1) is considerably greater than the recommended number for stable correlational analyses (typically, 250 to 500; Kretzschmar & Gignac, 2019, Schönbrodt & Perugini, 2013).

2) Figures/tables/data availability:
*I recommend adding a caption to the figure. Other than that, the tables are relevant and the data are available.*

We have now added a caption to Figure 1.

3) Ethical approval:
        *I did not notice a statement about ethics approval. I recommend adding one.*

We have now added an ethics statement in the procedure section clarifying that this research was approved by Northwestern University’s Institutional Review Board.

4) Language:
*The text is generally well written, but I recommend expanding and clarifying a few points about the methodology and results:

“It should be noted that this self-report assessment model was not supported by the evidence” (p. 8).*

We have now expanded on this point in the text.  *“This matrix was scored adjusting for overlapping keys and replacing the overlapping covariances with the corresponding best estimate of the item’s “true variance,” the average correlation for that item” (p. 9).*

We have clarified this statement and provided references in text for further reading on this approach.  *“Items were included as “best items" if they emerged for all ten iterations of k-fold cross validation” (p. 9).*

We now have expanded on this statement and provide references in text for further explanation on this method.  *p. 18. “Moreover, the structure of this self-informant agreement corroborates the overlapping structure of the 27 lower-level factors in the SPI.” How so?*

We have removed this sentence and clarified this point in multiple places in the paper. *Also, I think the Method section would be easier to follow if the Participants subsection were placed before the Analysis subsection.*

We have adjusted the order of the manuscript so that the Participants subsection is placed before the Analysis subsection.

**Editor Final Decision—Accept**

July 16, 2021

Dear Anne Zola,

I have now had a chance to read over your manuscript “The Convergence of Self and Informant Reports in a Large Online Sample”, along with the letter describing the changes you made in response to the editorial feedback. Thank you for your responsiveness to the concerns raised in the review process. I am happy to write 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. Our managing editor will contact you in case there are any pre-production file related questions. You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Have a nice weekend!

Sincerely,
Brent Donnellan